ОКТОБАР-ДЕЦЕМБАР, 2022.

ИНФОРМАТИЧКА ТЕОРИЈА III

ТРИЛОГИЈА ИНФОРМАТИЧКИХ ПРИЛОГА

РАСТКО ВУКОВИЋ
ПРОФ. МАТЕМАТИКЕ
Бања Лука, Република Српска

Растко Вуковић, октобар-децембар. 2022.
Информатичка Теорија III – трилогија информатичких прилога
http://rvukovic.net/knjige/IT3.pdf

Предговор

Ово је трећи непосредни наставак „Информатичке Теорије“[1] и „Информатичке Теорије II“[2]. Прва је садржавала наслове од „1. Неквадратна матрица канала“, до „43. Условне вероватноће“, а друга од „44. Сфере потенцијала“ до „67. Шредингерова мачка“. Нумерацију наслова само настављам. Све три скрипте радим на основу сећања и записа питања и одговора на Блогу[3] актуелном у вези са мојом теоријом информације, или других разговора са колегама.

Не очекујем да моје текстове иначе још неко озбиљно чита, бар не још задуго. Када ми се понеко од ауторитета науке обрати обично одговорим да су ови наводни радови предалеко отишли како би њиховој заједници били незанимљиви, да би ми интерес јавности убио интересовање за њих, ослабио инспирацију и покварио играчку. Овом приликом се извињавам посебно упорним којима сам „одбрусио“ да ствари нису добро разумели и да ја не пишем научне већ научно-фантастичне приче.

Стварни разлози су заправо „сређивање радног стола“, јер су ми важни моменти расути којекуда и има их толико да плачу за једном добром генералком.

Аутор, 2022.

[1] Информатичка Теорија I, https://archive.org/details/it_20220709/
[2] Информатичка Теорија II, https://archive.org/details/it-2_20220904
[3] Блог, http://rvukovic.net/blog/

Садржај

V Део: Алтернативе

Три теме сам издвојио из претходне две скрипте и грубо превео на енглески „Сила неизвесности“, „Фреквенција слова“, „Границе расподела“ (Uncertainty Force, Letter Frequency, Limit Distributions). Тако укратко препричане оне би остале подцењене, или би каснији уређивачи могли поверовати да њихов значај нисам добро разумео. Зато ћу овде направити још један искорак у том правцу.

Поменута „сила неизвесности“ у прилогу се позива на прву скрипту, на генерисање (компјутером) низова случајних исхода и запажање да такви, као вектори апстрактног простора вероватноћа, као и потег сунце-планета пребришу једнаке површине у једнаким временима, те да им се врх креће по хиперболи. На тај начин, за неизвесности важи Кеплеров други закон гравитације, или рецимо тачније, у основи свих константних централних сила лежи та законитост неизвесности. То откриће иде и даље у дефинисање потенцијала неизвесности, која формално одговара физичким.

Друга је скрипта започета анализом тог потенцијала помоћу једнаковероватних сфера, кругова и кружница са чијим полупречником расте извесност, а затим је дато и геометријско извођење као подршка претходним експериментима са насумичним. Подсећам да је ова, иначе свету науке нова идеја, заправо мени стара прича о кретању физичких набоја под утицајем опште спонтане тежње ка мање информативним стањима, дакле ка мањем дејству, сагласно чешћој реализацији више вероватних исхода. На пример, у књизи „Минимализам Информације“ у наслову „2.5 Ајнштајнове опште једначине“ приложио сам извођење једначина опште релативности на основу принципа најмањег дејства.

Други прилог, о фреквенцији слова, говори о истом на мало другачији начин. Јасан, прецизан, или садржајан говор мање је информативан! Организујући мисли ускраћујемо им опције, неизвесност им кратимо тако да Шенонова информација постаје мања од „чистог шума“, тј. потпуно насумично набацаних слова. Ту је такође на делу „сила неизвесности“ која чини да нас усмереност привлачи тачно онолико колико нас вишак могућности одбија. Свако биће (живо више од неживог једнаких категорија) поседује неку количину слободних опција, а закон одржања информација мири та два антагонизма (одбијања од веће неизвесности, односно привлачења мањој).

Из трећег прилога, који говори о ограничењима расподела вероватноћа, узео сам и наслов овога дела скрипти. У њему је једна важна и запостављена теорема (Theorem 12). Она каже да случајна промењљива на коначном домену постиже максималну Шенонову информацију са униформном расподелом, док ће са унапред задатом средњом вредношћу (математичким очекивањем) она то достићи у експоненцијалној расподели, а са задатом варијансом (дисперзијом) у нормалној. Оно што следи наставак је ова три прилога.

68. Опадање информације

Без журбе, прво приметимо да природа не воли једнакост (Equality) и ту тезу размотримо помоћу „мрежа без обима“ (Scale-free networks). Математичка теорија графова која их обрађује, као што је то иначе са математиком ствар, веома је универзална и практична. Замислимо да су чворишта тих мрежа рецимо имаоци моћи, новца, познанстава, или просто чворишта интернета, далековода, па и ћелије живог организма, а повезнице су оно што комуницира, или интерагује између чворишта. Осврнућемо се и на информатичке разлоге због којих је структура тих чистих „слободних мрежа“ заправо ретка (Scale-free networks are rare).

Приоритет је да су све повезнице, ма којих чворова дате мреже, међусобно равноправне. Ако тада захтевамо равноправност и чворова, такву да сваки има једнак број повезница, може се показати да мрежа није најефикаснија. На пример када би сваки човек имао једнако новца (моћи, утицаја, познаника) колање капитала (ефикасност, продорност, обавештеност) било би слабије, одлуке би биле краћег даха и спорије, односно економска моћ државе мања.

Понашање природе у деловању начелног шкртарења информацијом (комуникацијом, дејством) се види у спонтаном повећању тих мрежа. Додајући нове повезнице, чворови који имају бар по једну више од околних добиће нову повезницу са већим шансама од околних. Са још већим упоредним бројем повезница и вероватноћа добијања нових биће и већа. Равноправност чворова све се више нарушава, а мрежа стреми ка једном чвору који има све повезнице и свим осталим који их имају по једну. Међутим тај екстремни циљ опет је равноправност „свих осталих“ чворова, што спонтани развој догађаја управљан начелним шкртарењем природе информацијом неће дозволити.

Када направимо „топ листу“ чворова у мрежи који имају x веза са другим чворовима за велике вредности x, расподела вероватноћа у најпростијем случају показује степену законитост

$$P(x) = \beta x^{-\gamma}$$

где је γ експонент који зависи од мреже и обично је број између 2 и 3, а β константа нормирања. Ова вероватноћа спорије опада од експоненцијалне, па иако је информација степене расподеле у сваком случају мања од униформне, ипак остаје већа од експоненцијалне.

Оно што спречава даље опадање „количине неизвесности“ (информације) у развоју ових мрежа је избегавање превеликог равнања броја повезница осталих чворова, чија би равноправност почела расти. Продирање неизвесности ка зонама мање информације подсећа на прелевање воде према нижим нивоима. Или, то је модел сличан ширењу заразне болести која постепено ствара све већи број отпорних, а у случају смртности проређујући популацију, тако да у оба случаја она не достиже неко експоненцијално ширење.

Међутим, радиоактивно распад тешких атома нема наведену препреку. Усудићу се претпоставити да све честице природе, укључујући и оне за које данас сматрамо да се не распадају, а које су „без памћења“, имају експоненцијалну законитост расподеле свог распадања

$$P(t) = \alpha e^{-\beta t}$$

где је t време, β је константа зависна од врсте честица, а α је константа нормирања вероватноће. То шире виђење експоненцијалне расподеле [41. Ограничења] прилично се уклапа у физику ових процеса (Particle decay), иако не мислимо на исто. Супстанца памти и простор памти, углавном.

Корак од ове (хипо)тезе је тумачење гравитација помоћу експоненцијалне расподеле вероватноћа (Простор-Време, 1.2.8 Вертикалан пад), где смо видели да је

$$m = m_0 e^{GM/rc^2}.$$

Наиме, тело које пропада под утицајем гравитације планете масе M на удаљености r од центра гравитације, због повећања кинетичке енергије, $E_0 = m_0 c^2 \to E = mc^2$, добија на маси према законитости експоненцијалне расподеле. G је гравитациона константа, а c брзина светлости. Из наоко преједноставног извођења ове формуле следи иста зависност фреквенција, па све до саме метрике опште релативности.

Случајна променљива ове „расподеле вероватноћа“ је убрзање (Gravitational acceleration)

$$x = -\frac{G}{rc^2}, \quad f_e(x) = Me^{-Mx},$$

при чему масу планете M треба поделити јединичном масом ради добијања бездимензионих величина за потребе теорије вероватноће.

У време писања поменуте књиге (пре маја 2017.) нисам се усудио ићи толико ка данашњем ставу, о можда повезаности наведене формуле са експоненцијалном расподелом вероватноћа, све док нисам приметио и наводио извођење Ајнштајнових једначина гравитације из Ојлер-Лагранжових једначина. Познато је да се оне изводе принципом најмањег дејства – који сматрам еквивалентом минимализма информације. Та крајња идеја још увек је хипотеза, додуше „веома привлачна“.

Претходно запажање о сметњама тежњи ка експоненцијалној расподели важи и овде. У све јачим гравитационим пољима развија се све већи отклон, па гравитациона сила као да мења своју нарав што ће доследно теорији информације бити због простора који памти. Зато гравитација повлачи за собом то више простора што је јача, па путања Мекрура постаје елипса чији перихел ротира током планете, а масивније галаксије стварају „тамну материју“.

Настављајући приметимо да ће замена убрзања x брзином y, тела у слободном паду (Falling from Rest) на датом месту, експоненцијалну расподелу превести у нормалну

$$f_n(y) = \frac{e^{-\frac{1}{2}\left(\frac{y}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}},$$

где дисперзија σ зависи од масе M планете на коју тело пада.

69. Диференцијалне једначине

Претходна разматрања могуће је даље поопштавати. Наиме, када сматрамо да је информација у ткиву простора, времена и материје, а да је неизвесност њена бит, онда доследно све једначине физике треба размотрити као могуће једначине информатике. При томе можемо поћи од путања које су у познатој физици решења диференцијалних једначина најдаље другог реда.

Посматрајмо шта се догађа са збиром две експоненцијалне функције (C_1, C_2, a, b су константе):

$$y(x) = C_1 e^{-ax} + C_2 e^{-bx},$$

$$y' = -aC_1 e^{-ax} - bC_2 e^{-bx},$$

$$y'' = a^2 C_1 e^{-ax} + b^2 C_2 e^{-bx}.$$

Посматрајмо ове изводе као хомогени систем линеарних једначина са непознатима C_1 и C_2, које треба елиминисати:

$$\begin{vmatrix} y & -e^{-ax} & -e^{-bx} \\ y' & ae^{-ax} & be^{-bx} \\ y'' & -a^2 e^{-ax} & -b^2 e^{-bx} \end{vmatrix} = 0,$$

$$\begin{vmatrix} y & -1 & -1 \\ y' & a & b \\ y'' & -a^2 & -b^2 \end{vmatrix} = 0,$$

$$y \begin{vmatrix} a & b \\ -a^2 & -b^2 \end{vmatrix} - y' \begin{vmatrix} -1 & -1 \\ -a^2 & -b^2 \end{vmatrix} + y'' \begin{vmatrix} -1 & -1 \\ a & b \end{vmatrix} = 0,$$

$$(a-b)[y'' + (a+b)y' + ab] = 0.$$

Дакле, када $a - b \neq 0$, ово је диференцијална једначина другог реда (константних коефицијената) чије решење, полазну функцију $y = y(x)$, можемо одмах писати познајући њене коефицијенте. А коефицијенте диференцијалне једначине $y'' + py' + q = 0$ са коефицијентима њених решења a и b везују једнакости $p = a + b$ и $q = ab$, који су заправо решења (по r) квадратне једначине

$$r^2 + (a+b)r + ab = 0.$$

Важи и уопште, у математици, да линеарна једначина реда $n = 1,2,3, \dots$, има свој карактеристични полином чији корени (који су решења припадне полиномске једначине $r_1, r_2, \dots, r_n$, са некима који се понављају, или су комплексни бројеви) дефинишу експоненцијалне функције ($C_k e^{-r_k x}$, индекса $k = 1, 2, \dots, n$) чији збир даје решење диференцијалне једначине.

Пример 1. Када је $b = a$, тада је функција

$$y(x) = (C_1 x + C_2) e^{-ax}$$

решење горње диференцијалне једначине. То проверите непосредним уврштавањем. ❑

Пример 2. Када су решења карактеристичне једначине ($a, b = \alpha \pm i\beta$) коњуговано комплексни бројеви, тада решење диференцијалне једначине можемо поједноставити овако:

$$y(x) = C_1 e^{-(\alpha + i\beta)x} + C_2 e^{-(\alpha - i\beta)x} =$$

$$= C_1 e^{-\alpha x}(\cos\beta x + i\sin\beta x) + C_2 e^{-\alpha x}(\cos\beta x - i\sin\beta x)$$

$$= (C_1 + C_2)e^{-\alpha x}\cos\beta x + i(C_1 - C_2)e^{-\alpha x}\sin\beta x$$

$$= e^{-\alpha x}(A_1\cos\beta x + A_2\sin\beta x),$$

где су A_1, A_2, α, β константе. ❑

Тим познатим деловима математике сада додајемо (хипо)тезу о вероватносној природи решења диференцијалних једначина. Њихово виђење као егзактних (детерминистичких) појава долази из разлика микро и макро света бар по питању закона великих бројева теорије вероватноће. Овога ћемо се држати док (ако) се не покаже супротно.

Други додатак који доноси теорија информације је дубље разумевање комплексних бројева у физици. У 2. примеру, у случају да су C_1 и C_2 реални бројеви приметимо да је $A_2 = i(C_1 - C_2)$ имагинаран, или обратно, ако су први имагинарни други је реалан. Тачније и први и други могу бити комплексни бројеви, али ми у теорији диференцијалних једначина кажемо да су они неке произвољне константе. Ту слободу математике даље преносимо на физику.

Комплексна решења неких једначина физике, које класична теорија одбацује, сада третирамо као псеудо-реална, можда само привремено одсутна из реалности и присутна у њој „паралелној“ док се додатним операцијама имагинарности не пониште (Bypass). Одговарајући комплексни догађаји можда и даље негде постоје, али нам на датом месту и тренутку нису доступни. То су минијатурне варијације са становишта нашег макро света, због чега их олако одбацујемо.

Пример 3. Једначина $y = A\sin(\omega x + \varphi)$ представља синусоиду, или таласно кретање (x је време), где је ω угаона брзина, а φ фазни помак. Први, па други извод дају $y'' + \omega^2 y = 0$. То проверавамо диференцирањем и уврштавањем, или применом горњих формула. ❑

Овај пример показује колико је лако заметнути имагинарност решења диференцијалне једначине која представља таласно кретање. Наиме, из карактеристичне квадратне једначине

$$r^2 + \omega^2 = 0$$

добили бисмо њена решења $r_{1,2} = \pm i\omega$, а затим и решења одговарајуће диференцијалне:

$$y = C_1 e^{+i\omega x} + C_2 e^{-i\omega x} =$$

$$= (C_1 + C_2)\cos\omega x + i(C_1 - C_2)\cos\omega x$$

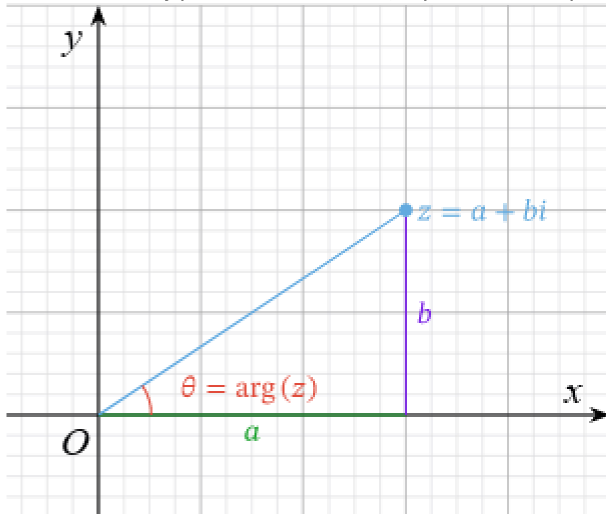
$$= A_1\cos\omega x + A_2\sin\omega x$$

$$= A\sin(\omega x + \varphi).$$

Приметите колико је пута у овим трансформацијама „гурнута под тепих“ имагинарност да бисмо на крају добили једну чисту реалност.

70. Комплексни простори

У комплексној равни $z \in \mathbb{C}$ апсцису називамо реалном осом $a = Re(z)$, а ординату имагинарном $b = Im(z)$, на слици лево. Угао од апсцисе до потега из исходишта O ка тачки $z = a + ib$ је $\theta = \arg(z)$. Дужина потега је $|z| = \sqrt{a^2 + b^2}$, број који називао модулом. Бројеви a и b су реални, а за број z кажемо да је комплексан, јер садржи реалан a и имагинарни сабирак ib. За имагинарну јединицу важи $i^2 = -1$.



Коњуговано комплексан број броја z такође је комплексан број $\bar{z} = a - ib$. Апсциса је оса симетрије та два, а њихов производ:

$$\bar{z}z = z\bar{z} = (a + ib)(a - ib) = |z|^2$$

је квадрат модула. Понављам добро познате појмове елементарне математике, да не буде забуне у каснијем позивању на њих (Trougao).

Када је дата реална функција $f(x)$, таква да је за сваку реални варијаблу x њена вредност реалан број, тада је $\bar{f}(z) = f(\bar{z})$. То је такође елементарна ствар, па је овде не доказујем. Приметимо да збир ових коњугованих, као и њихов производ, дају реалан број. Има и других начина да рачуном добијамо реалне бројеве радећи са комплексним. То је (један од разлога) да у резултатима где се имагинарни или комплексни бројеви појављују у физици (Solenoid) дајем псеудо-реална значења, а да их (мојој) „теорији информације“ чак третирам једнако са осталим решењима.

Пример 1. Са слике видимо да се иста тачка може писати помоћу косинуса и синуса:

$$z = |z|(\cos\theta + i\sin\theta) = |z|e^{i\theta},$$

јер је $a = |z|\cos\theta$ и $b = |z|\sin\theta$. Множећи овај са другим комплексним бројем:

$$zw = |z||w|(\cos\theta + i\sin\theta)(\cos\varphi + i\sin\varphi) =$$

$$= |z||w|[(\cos\theta\cos\varphi - \sin\theta\sin\varphi) + i(\sin\theta\cos\varphi + \cos\theta\sin\varphi)]$$

$$= |z||w|[\cos(\theta + \varphi) + i\sin(\theta + \varphi)]$$

$$= |z||w|e^{i(\theta+\varphi)}.$$

Дакле, производ комплексних бројева је комплексни број модула једнаког производу модула, а аргумента једнаког збиру аргумената фактора. ❑

Множењем комплексних бројева сабирају им се аргументи, а множењем једног са коњугованим другим аргументи се одузимају. То нам даје могућност да косинус и синус углова изразимо само помоћу производа комплексних бројева, као у следећем примеру.

Пример 2. За два произвољна комплексна броја $u = u_x + iu_y$ и $v = v_x + iv_y$ биће:

$$\bar{u}v = (u_x - iu_y)(v_x + iv_y) = (u_xv_x + u_yv_y) + i(u_xv_y - u_yv_x).$$

Коњугован овом производу је $\overline{\bar{u}v} = u\bar{v} = (u_x v_x + u_y v_y) - i(u_x v_y - u_y v_x)$. Углови које производи тих коњуговано комплексних бројева граде са апсицсом су $\pm\varphi = \pm\sphericalangle(u, v)$, где предзнак означава смер угла. Пројекције модула (интензитета) овог производа на координатне осе равни $\mathbb{C}$, реалну x и имагинарну y, налазимо за косинус и синус:

$$u_x v_x + u_y v_y = |u||v| \cos\varphi, \quad u_x v_y - u_y v_x = |u||v| \sin\varphi,$$

где је $|u| = \sqrt{u_x^2 + u_y^2}$, $|v| = \sqrt{v_x^2 + v_y^2}$ и $\varphi = \sphericalangle(u, v)$. ❑

Сагласност резултата из 1. примера, множења реалних и имагинарних експонената показује се и у једнакости $e^{i\theta}e^{i\varphi} = e^{i(\theta+\varphi)}$, а сагласност резулатата 2. примера са основним тригонометријским идентитетом $\cos^2\varphi + \sin^2\varphi = 1$ лако је проверити непосредним квадрирањем и сабирањем. Оно што ћемо даље видети је извођење формула ротације у комплексној равни.

Пример 3. Нека су $u = u_x + iu_y$ и $v = v_x + iv_y$ јединичних модула, $|u| = |v| = 1$. Тада је:

$$\begin{cases} v_x = (u_x^2 + u_y^2)v_x = (u_x^2 v_x + u_x u_y v_y) - (u_x u_y v_y - u_y^2 v_x) \\ v_y = (v_x^2 + v_y^2)v_y = (u_x^2 v_y - u_x u_y v_x) + (u_x u_y v_x + u_y^2 v_y) \end{cases} =$$

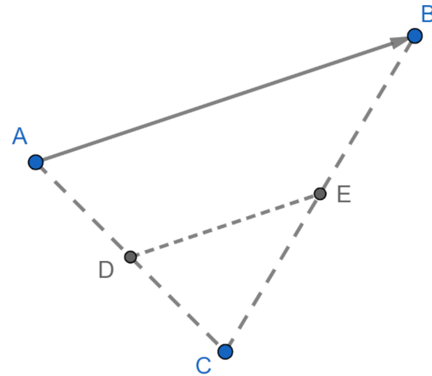
$$= \begin{cases} u_x \cdot \dfrac{u_x v_x + u_y v_y}{|u||v|} - u_y \cdot \dfrac{u_x v_y - u_y v_x}{|u||v|} = u_x \cos\varphi - u_y \sin\varphi \\ u_x \cdot \dfrac{u_x v_y - u_y v_x}{|u||v|} + u_y \cdot \dfrac{u_x v_x + u_y v_y}{|u||v|} = u_x \sin\varphi - u_y \cos\varphi \end{cases}$$

А то су формуле ротације $u \to v$ за угао φ око исходишта O. ❑

Резултат овог примера лакше препознајемо ако га напишемо матрично

$$\begin{pmatrix} v_x \\ v_y \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} u_x \\ u_y \end{pmatrix}.$$

Овако задата ротација важи и за произвољне али међусобно једнаке модуле, $|u| = |v|$. Ротацију за испружен угао $(180°)$ у равни називамо „централном симетријом“, ротација равни за угао $180°$ око њене осе (праве) је „осном симетријом“, ротација 3-дим простора око равни за $180°$ назива се огледалска или „равнинска симетрија“. Узастопне централне симетрије дефинишу „транслацију“, као што видимо на слици десно.



Тачке A и C симетричне су у односу на њима централну D, а C и B у односу на E. Дуж DE је средња линија троугла ABC и, према томе, паралелна је основици AB и има пола њене дужине. Када су тачке D и E фиксиране, две дужине DE су вектор транслације тачака тог простора, каква је транслација из тачке A у тачку B.

Свака изометријска (која чува удаљености тачака) трансформација може се добити помоћу самих ротација. Примењено то на физику, свака трансформација величина за које важи закон одржања је репрезентација ротација. Сматрам да отуда долази рецимо и израз $\Psi = ae^{i\varphi}$ за таласни вектор

слободне честице у квантној механици, где је a амплитуда а угао $\varphi = (\vec{p}\cdot\vec{r} - Et)/\hbar$. Посебно, два вектора $\vec{p} = (p_x, p_y, p_z)$ и $\vec{r} = (x, y, z)$ представљају импулс и положај честице, E је износ енергије, t тренутак, а $\hbar = h/2\pi$ је редукована Планкоква константа (Free particle).

Оно што је нама занимљиво у даљој примени поменуте „идеје ротације“ су примена теореме из наслова „41. Ограничења“ и уопштавање најављено у наслову „05. Хартлијева информација“, али отом потом.

70.1. Вектори

Да би боље разумели о чему говоре претходни примери, нарочито трећи, подсетимо се особина векторских простора (Квантна Механика). Геометријско представљање вектора са орјентисаним дужима је једно од најједноставнијих и најпознатијих. У горе приказаној комплексној равни $z \in \mathbb{C}$ апсцису и ординату дефинишу узајамно окомити јединични вектори, рецимо ознака $\vec{e}_x$ и $\vec{e}_y$. Тада комплексни број постаје вектор $\vec{z} = \vec{a} + \vec{b} = a\vec{e}_x + b\vec{e}_y = (a, b)$. Наведени су равноправни записи истог вектора.

Скаларни производ вектора, рецимо $\vec{u} = (u_x, u_y)$ и $\vec{v} = (v_x, v_y)$, је производ њихових интензитета и косинуса угла између $\vec{u}\cdot\vec{v} = |\vec{u}||\vec{v}|\cos\varphi$. Интензитет векторског производа вектора је производ њихових интензитета и синуса угла између $|\vec{u}\times\vec{v}| = |\vec{u}||\vec{v}||\sin\varphi|$, иначе вектор окомит на раван у којој леже дата два, осе јединичног вектора $\vec{e}_z$. На тај начин множења даће:

$$\vec{e}_x\cdot\vec{e}_x = 1, \quad \vec{e}_x\cdot\vec{e}_y = 0, \quad \vec{e}_y\cdot\vec{e}_x = 0, \quad \vec{e}_y\cdot\vec{e}_y = 1,$$

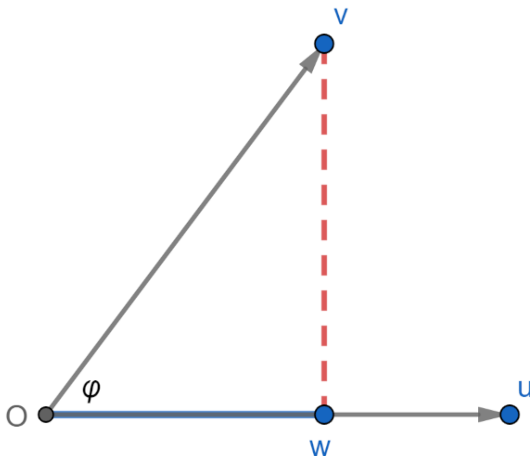
$$\vec{e}_x\times\vec{e}_x = 0, \quad \vec{e}_x\times\vec{e}_y = \vec{e}_z, \quad \vec{e}_y\times\vec{e}_x = -\vec{e}_z, \quad \vec{e}_y\times\vec{e}_y = 0.$$

Затим, множећи дате векторе по компонентама, налазимо:

$$\vec{u}\cdot\vec{v} = (u_x\vec{e}_x + u_y\vec{e}_y)\cdot(v_x\vec{e}_x + v_y\vec{e}_y) = u_xv_x + u_yv_y = |\vec{u}||\vec{v}|\cos\varphi,$$

$$\vec{u}\times\vec{v} = (u_x\vec{e}_x + u_y\vec{e}_y)\times(v_x\vec{e}_x + v_y\vec{e}_y) = (u_xv_y - u_yv_x)\vec{e}_z = |\vec{u}||\vec{v}|\sin\varphi\ \vec{e}_z.$$

На следећој слици лево приказани су ови вектори и окомита пројекција другог на први. Видимо



правоугли троугао Owv са катетама:

$$|Ow| = |Ov|\cos\varphi, \quad |wv| = |Ov|\sin\varphi.$$

То је и одговор на горње питање, да $u_xv_x + u_yv_y$ представља доњу катету, а $u_xv_y - u_yv_x$ горњу, а множене још само интензитетом доњег вектора.

Када су вектори $\vec{u}$ и $\vec{v}$ јединични, они се налазе на јединичној кружници са центром у O, што значи на тригонометријској кружници истог исходишта, а поменуте катете буквално су косинус и синус тог угла (φ).

Међутим, вектори нису само „орјентисане дужи“, па ова тема тражи још пажње.

Вектори су и матрице. Да за матрична множења не важи закон комутације видимо у примеру.

Пример 4. Дате су матрице другог реда:

$$\hat{A} = \begin{pmatrix} a + \frac{1}{4} & 1 \\ 1 & a - \frac{1}{4} \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} b - \frac{1}{4} & 1 \\ 1 & b + \frac{1}{4} \end{pmatrix}.$$

Нађимо њихове комутиране производе:

$$\hat{A}\hat{B} = \begin{pmatrix} ab - \frac{1}{4}(a-b) - \left(\frac{1}{4}\right)^2 + 1 & a + b + \frac{1}{2} \\ a + b - \frac{1}{2} & 1 + ab + \frac{1}{4}(a-b) - \left(\frac{1}{4}\right)^2 \end{pmatrix},$$

$$\hat{B}\hat{A} = \begin{pmatrix} ab - \frac{1}{4}(a-b) - \left(\frac{1}{4}\right)^2 + 1 & a + b - \frac{1}{2} \\ a + b + \frac{1}{2} & 1 + ab + \frac{1}{4}(a-b) - \left(\frac{1}{4}\right)^2 \end{pmatrix}.$$

Ови резултати нису једнаки, проверавамо одузимањем:

$$\hat{A}\hat{B} - \hat{B}\hat{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Некомутативност није увек константна, као у овом случају. ❑

Резултат примера је друга матрица кватерниона (Quaternion fields) другог реда. Писано помоћу тих кватерниона и комутатором:

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \hat{q}_2.$$

Квадрати кватерниона су негативне јединичне матрице, па су они у том смислу проширење појма имагинарног броја. У случају матрица другог реда то су три кватерниона:

$$\hat{Q}_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \hat{Q}_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \hat{Q}_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

За сваку од ових важи $\hat{Q}^2 = -\hat{I}$, што је лако проверавати непосредним множењем. Аналогије броју један су Паулијеве матрице:

$$\hat{P}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{P}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{P}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Непосредним множењем лако се проверава да за сваку од ових важи $\hat{P}^2 = \hat{I}$. На тај начин, као у прилогу датог линка, успевао сам радити са 6-дим „простор-временом“, из којег се могу узети ма које четири координате, три просторним x, y, z, а четврта имагинарном временском дужином ict. То је било пре више од деценије, али због моје склоности анонимности, за ширу јавност би та иста прича била још увек нова.

Нагласићу зато још једном, једначине опште релативности, као и многе квантно-механичке, веома се добро понашају у моделу 4 димензије простор-времена произвољно извучених из 6 димензија наводног простор-времена моје теорије информације.

Линеарни оператори су такође вектори. У следећем примеру су два таква физике.

Пример 5. Оператори импулса и положаја су $\hat{p} = -i\hbar\frac{d}{dx}$ и $\hat{x} = x$ квантне механике. Они делују на таласну функцију на начине:

$$\hat{p}\psi = i\hbar\frac{d\psi}{dx}, \quad \hat{x}\psi = x\psi,$$

$$[\hat{p},\hat{x}]\psi = (\hat{p}\hat{x} - \hat{x}\hat{p})\psi = \hat{p}\hat{x}\psi - \hat{x}\hat{p}\psi =$$

$$= -i\hbar\frac{\partial}{\partial x}(x\psi) + xi\hbar\frac{\partial}{\partial x}\psi = -i\hbar\left(\psi + x\frac{\partial\psi}{\partial x}\right) + i\hbar x\frac{\partial\psi}{\partial x} = -i\hbar\psi,$$

$$[\hat{p},\hat{x}] = -i\hbar.$$

Ова једнакост је основа релација неодређености. ❏

Репрезентације неких линеарних оператора (матрица) су процеси квантне механике, а вектора на које ови делују, су стања. Сходно томе, некомутативност оператора казује да је редослед радњи у микро процесима битан. Степен некомутативности исказан комутатором дефинише неодређеност код првог процеса када делујемо другим и обрнуто. Константна неодређеност из горњег примера, реда величине ħ, је Хајзенбергова неодређеност импулса и положаја честице-таласа, а на тај исти начин дејства енергије и времена. При томе су имагинарности доследне теорији информације.

70.2. Квантна стања

Оператори су промене на стањима, а квантна стања су нарочите врсте вектора, таласних функција. Међу најпростије случајеве квантних стања спадају слободне честице:

$$\psi(\vec{r},t) = Ae^{i(\vec{k}\cdot\vec{r}-\omega t)} = Ae^{i(\vec{p}\cdot\vec{r}-Et)/\hbar},$$

где је A амплитуда таласа (мерљивост), $\vec{p}$ је импулс таласа, $\vec{k}$ је таласни вектор (број), $\vec{r}$ положај, E енергија, t време.

Израз у загради експонента представља имагинарни угао $\varphi = (\vec{p}\cdot\vec{r} - Et)/\hbar$ поменутих „ротација“ таласа-честице. Ако је угао $\varphi = \pm n\pi$ у радијанима, редом за $n = 0, 1, 2, \dots$, биће $e^{i\varphi} = \pm 1$, а у свим осталим случајевима $e^{i\varphi}$ је имагинаран број. Ми га у оба случаја третирајмо као експоненцијалну функцију и, доследно претходном излагању, сматрајмо да ова експоненцијална функција одређује неку „вероватноћу“, а њен логаритам одговарајућу „информацију“. Квантна механика већ познаје такву интерпретацију у Борновом закону, а овде чинимо још само један корак ка „заобилажењу“.

Какав год је „угао ротације“ φ, таласну функцију ψ држимо само за половину решења, обзервабле (физички мерљиве величине), док је стварна густина вероватноће налажења честице таласа:

$$\rho(\vec{r},t) = \overline{\psi}(\vec{r},t)\psi(\vec{r},t) = |\psi(\vec{r},t)|^2,$$

где се уместо ознаке коњуговања $\overline{\psi}$ уобичајено користи и ознака ψ^*. Поред овога треба знати да се таласна функција слободне честице често представља као суперпозиција сопствених функција момента, са коефицијентима датим Фуријеовом трансформацијом почетне таласне функције. О томе сам писао раније, популарно (Приче о информацији), или математички (Квантна Механика, 1.3.7 Ортогоналност), а можда ће опет бити прилика.

Практично је уобичајена појава таласног пакета, који путује као јединица. Пакет таласа сматрамо бесконачним скупом компонентних синусоидних таласа различитих таласних бројева, али се може анализирати и парцијално, у паровима сабирака попут:

$$u(x,t) = Ae^{i(kx-\omega t)} + Be^{-i(kx-\omega t)},$$

где је претходни таласни вектор $\vec{k} = (k_x, k_y, k_z)$ разложен у таласне бројеве k, а све три просторне димензије сведене (апроксимиране) на аспцису. Свака оваква компонентна таласна функција, а са тиме и таласни пакет, решења су (Шредингерове) таласне једначине.

Доследно „теорији информације“, овде приметимо да се сваки линеарни оператор да раставити на факторе (разложити на композицију) линеарних оператора, посебно и сваки унитарни оператор (њихове су репрезентације процеси квантне механике) може се такође писати као производ оних унитарних оператора који представљају (под)процесе датим. То иде у прилог оном делу тумачења (моје) теорије где се држи да неизвесност природа извлачи и из бесконачности, те да је коначност само део реалности.

На пример, увек можемо писати:

$$\psi(\vec{r},t) = Ae^{i(\vec{k}\cdot\vec{r}-\omega t)} = A_1A_2e^{i(\overrightarrow{k'}\cdot\overrightarrow{r'}-\omega' t')}e^{i(\overrightarrow{k"}\cdot\overrightarrow{r"}-\omega" t")} = \psi'\psi",$$

што је алгебарски сасвим коректно, а сада имамо и информатичке интерпретације. Наиме, ма који скаларни производ може се разложити на два скаларна производа

$$\vec{k}\cdot\vec{r} = \vec{k}'\cdot\vec{r}' + \vec{k}''\cdot\vec{r}''$$

као што се сваки број може писати као збир два броја, а исто је са временским додатком ωt. Онда је даље тачно и разлагање експонената, $e^{a+b} = e^a e^b$. То није спорно. Спорно је тумачење, да ови недељиви делови, које теорија информације такође познаје (Packages), имају неке физици слабије видљиве састојке са својствима који би се могли показати не противречним (математичким), или ће бити разлог ревизије неких ставова данашње науке.

70.3. Спрега

Ова разлагања таласних функција, суперпозиције и растављања на факторе, воде нас директно ка облицима „информације перцепције“. То су заправо рачунски поступци присутни одавно, од самог настанка квантне механике, али као да су остали непопуларни због тада жестоких критика[4]. Они су део теме квантне спрегнутости (EPR paradox) до данас више пута експериментално потврђене, али чије се непосредне теоријске основе још увек покушавају избегавати. Погледајмо део тог чувеног

[4] A. Einstein, B. Podolski and N. Rosen: Can Quantum-Mechanical Description or Physical Reality Be Considered Complete? https://cds.cern.ch/record/405662/files/PhysRev.47.777.pdf

Ајнштајн-Подолски-Розеновог рада из 1935. године, којим откривају квантну спрегнутост желећи негирати ваљаност математике квантне механике.

Нека су $a_1, a_2, a_3, \ldots$ својствене вредности неке физичке величине A које се односе на систем I и $u_1(x_1), u_2(x_1), u_3(x_1), \ldots$ одговарајуће својствене функције, где x_1 стоји за променљиве коришћене за опис првог система. Тада ψ, сматрано функцијом од x_1, можемо писати

$$\psi(x_1, x_2) = \sum_{n=1}^{\infty} \psi_n(x_2) u_n(x_1)$$

где x_2 стоји за променљиве коришћене за опис другог система. Овде $\psi_n(x_2)$ представља коефицијенте развоја ψ у ред ортогоналних функција $u_n(x_1)$. Нека је величина A сада мерена и нађена јој је вредност a_k. Тада је закључено да након мерења први систем остаје у стању дато таласном функцијом $u_k(x_1)$, а да је други систем остављен у стању представљеном таласном функцијом $\psi_k(x_2)$. То је процес редукције таласног пакета; таласни пакет дат горе наведеним бесконачним редом редукован је на један сабирак $\psi_k(x_2)u_k(x_1)$.

Скуп функција $u_n(x_1)$ одређен је избором физичке величине A. Да смо уместо ове, бирали другу величину, рецимо B, која има својствене вредности $b_1, b_2, b_3, \ldots$ и својствене функције $v_1(x_1), v_2(x_1), v_3(x_1), \ldots$, требали би уместо претходног збира имати

$$\psi(x_1, x_2) = \sum_{s=1}^{\infty} \varphi_s(x_2) v_s(x_1)$$

где су φ_s нови коефицијенти. Ако је тада мерена величина B и њена вредност била b_r, закључак би био да се после мерења први систем налази у стању $v_r(x_1)$ а други систем у $\varphi_r(x_2)$.

Даље се у том тексту примећује да последица два различита мерења на првом систему може бити да други остаје у стањима различитих таласних функција. А како време мерења два система није у међусобној интеракцији, сматрају даље, није могућа промена другог система као последица било чега што би могло утицати на први систем. Међутим, аутори превиђају могућност истовремености неког система док га други опажају у различитим временима, иако је таква ситуација веома важно полазиште Ајнштајнове теорије релативности.

Ту је поента објашњења те квантне спрегнутости моје „теорије информације“. Износим је одавно, али прилично анонимно, па има смисла још једном поновити. Посматрајмо виртуелну сферу која настаје у средишту електрона и шири се као фотон. Говорим о Фајнмановим дијаграмима, али на начин ревидиран „теоријом информације“. У сваком кораку ширења та је сфера истовремена за себе, а за друге посматраче различито-времена. Као последица те различитости, њена евентуална интеракција са још неким набојем произвешће њено сопствено истовремено „пражњење“ које ће другима изгледати као да се дешава у различитим временима. Оно им се дешава и на различитим местима, па ће остали учесници догађај видети као „фантомско деловање на даљину“.

Користим Ајнштајнов израз којим је описао не-локалне промене спрегнутих система, мислећи на два тако удаљена догађаја да светлост не би могла стићи од једног до другог. Када два спрегнута

догађаја не би могла бити истовремена (у сопственом систему), тада би „истовремена“ реакција другог инцирана променом првог заиста била „фантомска“.

Слично је са две честице (Алиса и Боб) које настају распадањем и у истом се тренутку крећу на две супротне стране, рецимо збирно нултог импулса и спина. Оне ће представљати један истовремен систем за себе, али не и за остале. Интеракција једне одразиће се истовремено на другој, иако ће за треће посматраче ове честице тада можда бити предалеко за међусобну размену светлосних сигнала (Quantum entanglement).

Описао сам, а надам се и објаснио, једну врсту спрегнутости квантних система који још увек уносе можда и највише нејасноћа у своју област физике. Али те повезаности нису све оно што је битно за информацију перцепције, нити су можда најбитније. Постоје многе друге занимљиве интеракције малих система које се не дешавају истовремено, а такође обухватају исте „дуалне спреге“. Све те интеракције су комуникације, скаларни производи, па и информације перцепције.

70.4. Фуријеови коефицијенти

Простор вектора (низова) комплексних бројева у којем је дефинисан скаларни производ назива се унитарни простор. Простор $C_{[-\pi,\pi]}$ непрекидних функција на интервалу $[-\pi,\pi]$ је унитаран јер има дефинисан скаларни производ

$$\langle f,g\rangle = \int_{-\pi}^{\pi} f(t)\bar{g}(t)\,dt$$

где се уобичајено пише $\langle f,g\rangle$ уместо $f\cdot g$. У том простору функције

$$e_k(t) = \frac{e^{ikt}}{\sqrt{2\pi}},\quad k = 0,\pm1,\pm2,\dots$$

образују ортонормиран систем вектора, јер је производ произвољне две:

$$\langle e_k,e_n\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(k-n)t}\,dt = \begin{cases}1, & k=n\\ 0, & k\neq n\end{cases}$$

За произвољну функцију f из наведеног простора број

$$\varphi_k = \langle f,e_k\rangle = \frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi} f(t)e^{ikt}\,dt$$

назива се Фуријеовим коефицијентом (Fourier series), а

$$S(f) = \sum_{k=-\infty}^{\infty} \varphi_k e^{ikt}$$

назива се Фуријеов ред функције f.

Кошијев низ вектора (функција, тачака) x_k је низ чији елементи постају произвољно блиски један другом како секвенца напредује. Другим речима, низ је Кошијев ако за било коју малу позитивну

унапред дату удаљеност $\varepsilon > 0$, су сви сем коначног броја елемената низа мање удаљени један од другога од те удаљености. Прецизније:

$$(\forall\varepsilon > 0)\ (\exists n_0 \in \mathbb{N})\ (\forall m, n > n_0)\ |x_m - x_n| < \varepsilon.$$

Простор у којем нема празнина, у коме сваки Кошијев низ конвергира, назива се потпун, односно комплетан простор.

Пример 6. Нека је x_n број добијен заокруживањем корена броја два ($\sqrt{2} = 1.414213562\ldots$) на n децимала. Сви чланови низа тако заокружених бројева рационални су бројеви, он је Кошијев, јер конвергира, али број којем он конвергира ($x_n \to \sqrt{2}$, када $n \to \infty$) није рационалан. У том смислу простор рационалних бројева $\mathbb{Q}$ има празнине, а простор реалних бројева $\mathbb{R}$ је потпун. ❑

Потпун унитаран простор зове се Хилбертов простор, а потпун нормиран простор Банахов простор. Ове дефиниције наводим ради лакше даље орјентације, јер нема потребе да понављам детаље и доказе из мојих ранијих текстова (нпр. Квантна Механика). Показаћу да делимични низ $s_n(x)$, који садржи првих n чланова $S(f)$, Фуријеовог реда функције f, конвергира ка f.

Пример 7. За сваки низ бројева $b_1, b_2, \ldots, b_n$ биће

$$0 \le \left| f - \sum_{k=1}^{n} b_k e_k \right|^2 = \langle f - \sum_{k=1}^{n} b_k e_k\, , f - \sum_{k=1}^{n} b_k e_k \rangle =$$

$$= \langle f, f \rangle - \sum_{k=1}^{n} b_k \langle e_k, f \rangle - \sum_{j=1}^{n} \bar{b}_j \langle f, e_j \rangle + \sum_{k,j=1}^{n} b_k \bar{b}_j \langle e_k, e_j \rangle.$$

где је $e_1, e_2, \ldots, e_n$ ортонормирана база датог простора. Отуда

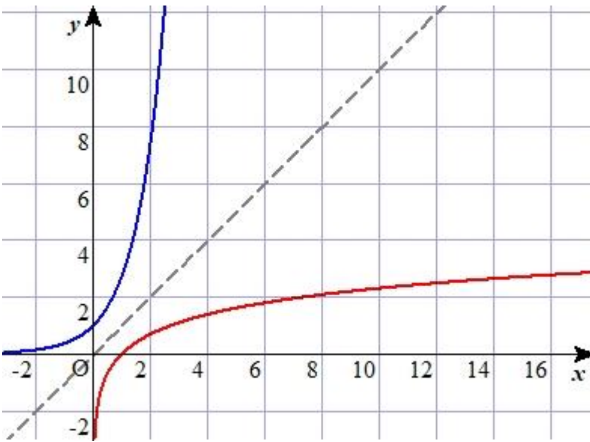
$$\left| f - \sum_{k=1}^{n} b_k e_k \right|^2 = |f|^2 - \sum_{k=1}^{n} |\varphi_k|^2 + \sum_{k=1}^{n} |b_k - \varphi_k|^2 \ge 0.$$

Трином у средини има најмању вредност ако и само ако је $b_k = \varphi_k$. Дакле најбоља апроксимација вектора (функције) f помоћу линеарног споја вектора $e_1, e_2, \ldots, e_n$ постиже се помоћу Фуријеових коефицијената. ❑

71. Логаритам

Функција је релација између елемената два скупа, домена и кодомена, при чему елеменат првог може бити придружен највише једном елементу другог. Тако, функција квадрирања $f(x) = x^2$ за једну вредност промењљиве x из домена неће бити придружена двема вредностима f кодомена. Обрнуто је функцијама дозвољено. На пример, $f(\pm 2) = 4$, што значи да се сваки од бројева плус и минус два може пресликати у по само један број (четири), али је обрнуто исти резултат могуће добити том функцијом пресликавањем два различита броја.

Овде је важно да приметимо да ће инверзно придруживање (пресликавање) функције опет моћи бити нека функција ако и само ако је полазна функција била „бијекција“, обострано једнозначно пресликавање, односно таква која је за сваки оригинал могла давати само по једну копију. Даље је разумљиво да ће функције и њима инверзне функције, ако такве постоје, представљене графом у Декартовом правоуглом систему кориднатa, бити осно симетричне у односу на симетралу I и III квадранта (праву $y = x$).



На слици лево, плаво је граф експоненцијалне функције $f(x) = e^x$, а црвене боје је граф њој инверзне логаритамске $f^{-1}(x) = \ln x$. Они се осно пресликавају преко испрекидане линије чија је једначина y$(x) = x$, која је самој себи инверзна функција.

Експоненцијална и логаритамска функција су бијекције и једна другој су инверзне, па је:

$$e^{\ln x} = x, \quad \ln e^x = x.$$

Отуда, из:

$$x_1x_2 = e^{\ln x_1}e^{\ln x_2} = e^{\ln x_1 + \ln x_2},$$

а са друге стране из

$$x_1x_2 = e^{\ln(x_1x_2)},$$

следи

$$\ln(x_1x_2) = \ln x_1 + \ln x_2$$

када год су ови логаритми дефинисани. Такву „адитивност“:

$$f(x_1)(x_2) = f(x_1 + x_2), \quad f^{-1}(x_1x_2) = f^{-1}(x_1) + f^{-1}(x_2),$$

имају само ове две функције, експоненцијална и логаритамска.

Са слике се види да је плави граф сав изнад апсцисе, а црвени десно од ординате. Другим речима кодомен експоненцијалне функције је позитиван број и управо зато је домен логаритма позитиван аргумент. Степеновање јединице (због $1^x = 1$) не сматрамо експоненцијалном функцијом, па тако ни логаритам базе 1 инверзном. Међутим, свака база дозвољена експоненцијалној ($b > 0$ и $b \neq 1$) дозвољена је и логаритамској функцији; ако је $f(x) = b^x$ онда је $f^{-1}(x) = \log_b x$.

Како из $b^x = y$ следи $\log_b y = x$, тако из $b^{nx} = y^n$ следи $\log_{b^n} y^n = x$, што значи да степеновање истим бројем базе и нумеруса не мења вредност логаритма. Доследно томе:

$$\log_b a^n = n \log_b a, \quad \log_{b^n} a = \frac{1}{n} \log_b a,$$

када су сви логаритми у изразима дефинисани, а отуда:

$$\log_b a = \log_b c^{\log_c a} = (\log_c a) \log_b c$$

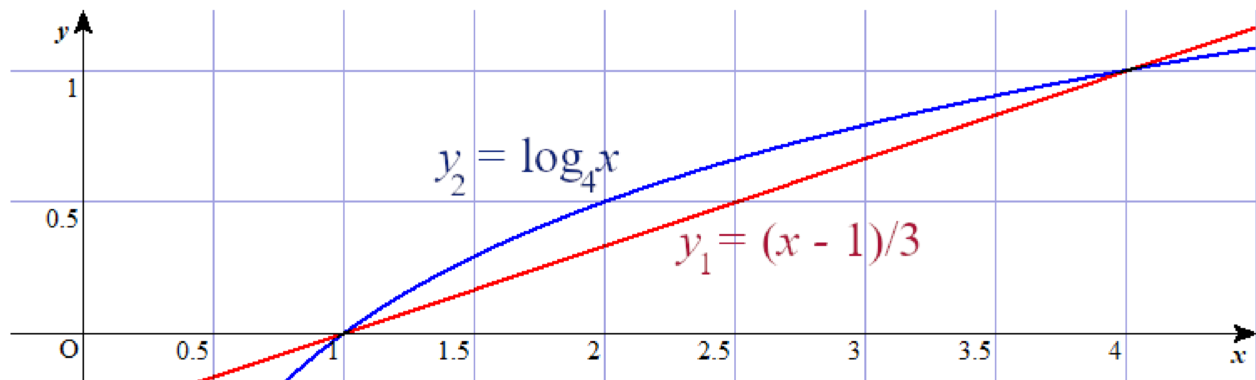
и правила за промену базе логаритма, на пример:

$$\log_b a = \frac{\ln a}{\ln b}, \quad \log_b a = \frac{1}{\log_a b}.$$

У другој једнакости примењена је $\log_b b = 1$.

71.1. Апроксимација

Логаритамску функцију $y_2(x) = \log_b x$, у крајњим тачкама интервала $[1, b]$ и унутар тог интервала, можемо апроксимирати правом $y_1(x) = (x-1)/(b-1)$, као што се види на слици за $b = 4$.



Највећа разлика $f(x) = y_2(x) - y_1(x)$ је у стационарној тачки коју налазимо помоћу извода:

$$f(x) = y_2 - y_1 = \log_b x - \frac{x-1}{b-1},$$

$$f'(x_0) = \frac{1}{x_0 \ln b} - \frac{1}{b-1} = 0,$$

$$x_0 = \frac{b-1}{\ln b},$$

а максимална разлика је број $f(x_0)$. Овај разломак (x_0) је растућа функција базе, а случају базе $b = 4$, на слици, апсциса и ордината екстрема су $x_0 \approx 2{,}164$ и $f(x_0) \approx 0{,}169$. Као што се може погодити са те слике, заједничке тачке ових функција иначе су $A(1{,}0)$ и $B(b, 1)$.

Развоји следећих функција у Маклоренове редове:

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots,$$

$$\ln(1 \pm x) = \pm x - \frac{x^2}{2} \pm \frac{x^3}{3} - \cdots - (-1)^n \frac{x^n}{n} - \cdots,$$

конвергирају за $|x| < 1$. За још мање варијабле ($x \to 0$) првих неколико чланова редова могу бити добре приближне замене самих функција.

71.2. Информација

У функцији $y = e^x$, односно $x = \ln y$, варијаблу x у интерпретација дефинише информацију, а y вероватноћу. То примећује Веберов (1834) закон опажаја[5], а експлицитно је изрекао Хатли скоро век касније.

Пример 1. У декадном систему, базе $b = 10$, до N-то цифрених бројева је 10^N. Хартлијева (1928) дефиниција информације каже да је N информација таквог скупа са 10^N једнако вероватних, или уопште, N је информација скупа са b^N једнако вероватних елемената изражена јединицама b (bit када је $b = 2$, nat када је $b = e$, decit када је $b = 10$). ❑

Применимо сада Хартлијеву формулу, да информација M једнако вероватних могућности износи $\log M$ у одговарајућим јединицама информације и покушајмо је даље применити на комплексни логаритам.

Ојлерова експоненцијална функција ($i^2 = -1$)

$$e^{i\theta} = \cos\theta + i\sin\theta$$

бијекција је за углове θ унутар интервала 0 до 2π радијана, такође и од 2π до 4π, заправо унутар сваког интервала дужине периоде $T = 2\pi$, при чему у том интервалу могу бити негативни углови. Комплексни број и његов логаритам, у поларном систему координата $Or\theta$, удаљености r тачке од исходишта O и угла θ између апсцисе и правца тачке, су:

$$z = re^{i\theta}, \quad \mathrm{Log}\, z = \ln r + i\theta.$$

То важи за сваки интервал периоде, па су ове функције периодичне, а отуда су и величине које се њима представљају периодичне. Комплексни логаритам је према томе генерализација природног логаритма на комплексне бројеве различите од нуле. Сматраћемо даље да бројеви z и $\mathrm{Log}\, z$ могу бити пар који дефинише „вероватноћу“ и одговарајућу „информацију“. Обзиром на значење овог угла θ у квантној физици, његову пропорционалност са производом импулса и положаја честице-таласа, односно енергије и времена, ову информацију сматрамо еквиваленту дејству.

Квантна физика за сада тежиште ставља на вероватнћама, дакле на експоненцијалним функцијама стања и процеса. Само на корак од њих је концепт информације о којој говоримо, бар онај део из непосредног ширења Хартлијеве информације. Промена такве

$$\frac{\partial\, \mathrm{Log}\, z}{\partial t} = \frac{\dot{r}}{r} + i\dot{\theta}$$

зависна је не само од реалне вероватноће (r), већ и од отклона у имагинарност (θ), што је можда разлог њеног слабог виђења у науци данас.

Сабирање независних вероватноћа квантних стања (честица-таласа), или „квантна суперпозиција“, основни је принцип квантне механике. Слично таласима класичне физике, квантна стања се могу

[5] Информација Перцепције, 1.6 Експерименталне потврде

сабирати („суперпоновати“) и резултат ће бити још једно важеће квантно стање; и обрнуто, да се свако квантно стање може представити као збир два или више других стања. Математички, то је својство решења Шредингерове једначине. Пошто је она линеарна, свака линеарна комбинација биће такође њено решење.

Међутим, експоненцијална и логаритамска функција нису узајамно линеарне, па онда то нису ни вероватноћа и информација у свом најпростијем облику. Зато и даље остаје практично решавање Шредингерове једначине за вероватноће обзервабли, а затим дискусија информације. На пример, иако су експоненцијалне расподеле $\alpha e^{-\alpha x}$ и $\beta e^{-\beta x}$ за $\alpha, \beta > 0$ добро нормиране, њихов производ $\alpha\beta e^{-(\alpha+\beta)x}$ није. Зато „информацију перцепције“

$$S = \alpha_1\beta_1 e^{-(\alpha_1+\beta_1)x} + \alpha_2\beta_2 e^{-(\alpha_2+\beta_2)x} + \cdots,$$

чак и када представља производе низова расподела, можемо посматрати као посебну величину.

71.3. Сила

Пример 2. Сила опруге са утегом масе m на њеном крају је

$$\vec{F} = -k\vec{x},$$

где је k константа опруге, а $x(t)$ је елонгација, или помак тела из равнотежног положаја. Њутнов други закон $F = ma$ даје диференцијалну једначину титрања:

$$m\ddot{x} + kx = 0,$$

са хармонијском функцијом као решењем [69.]

$$x(t) = A\cos\left(\frac{2\pi t}{T} + \varphi\right),$$

где је t време, а A, T, φ су константе. Прва је амплитуда, друга је период, при чему је $f = 1/T$ број титраја у јединици времена (фреквенција), а угао φ је фазни помак. ❑

У примеру видимо решење (хармонијску функцију) једног веома реалног догађаја, испред којег је следио израчунавање помоћу комплексних бројева, па њихов нестанак у коначном резултату. То је нешто слично „заобилажењу“ (Bypass), где „стварна“ честица одлази и у псеудо-стварна стања да би се поново појавила. За разлику од микро стања у којима се случајности могу боље виђати, у макро стањима доминирају закони великих бројева и оне, као и јасне преломне ситуације теорије хаоса, ретке су. Ми смо зато превише навикли да живимо без таквих, па лако поверујемо да нити њих, нити одлазака у „комплексни простор“ реалних честица уопште нема.

Оператор времена, $\partial_t x = \frac{\partial x}{\partial t} = \dot{x}$, тако делује да из својствене једначине $\dot{x} = -\lambda x$ следи експоненцијална зависност:

$$\frac{dx}{dt} = -\lambda x, \quad \frac{dx}{x} = -\lambda dt, \quad \ln x = -\lambda t + C_1, \quad x = C_2 e^{-\lambda t},$$

где је $C_2 = \exp C_1$ константа. Ове сам зависности анализирао раније у ИТ скриптама нарочито за случајеве параметра $\lambda > 0$. Тада смо могли видети, када је реч о расподелама вероватноћа, да се ове односе на канале преноса информација који не памте. То су ситуације попут бацања новчића

када следећи исход не зависи од претходних, па до честице-таласа квантне механике која на путу не мари за своја претходна стања и увек је „као нова“ (доследно закону одржања).

У истом примеру поучна је и сама диференцијална једначина $\ddot{x} = \lambda x$, где је $\lambda = -k/m$. Она нам каже да убрзање ($\ddot{x}$) опруге никада не мирује (осим у самој тачки промене смера, у исходишту), да је веће за веће истезање опруге (x) и сваки је пут усмерено ка равнотежној тачки (O). То је онај део хармонијске једначине који лако поопштавамо, рецимо помоћу „сила неизвесности“ на начине који их и интуитивно описују.

На пример, виртуелне сфере [38.], које иначе не треба посматрати само као виртуелне фотоне из можда Фајнманових дијаграма, већ уопште као центре извесности од којих са удаљеношћу расте неизвесност. Пренесена у „стварност“, таква апстракција постаје модел хармонијског осцилатора. Силу која гура опругу и производи титрање сводимо на вероватноће и начелни минимализам из (моје) теорије информације. Сличан је и следећи пример.

71.4. Својствене величине

Пример 3. Временски независна Шредингерова нерелативистичка једначина

$$\hat{H}\psi = E\psi,$$

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right]\psi(r) = E\psi(r).$$

На левој страни једнакости је Хамилтонијан који садржи набла оператор

$$\nabla^2\psi = \frac{\partial^2\psi}{\partial x^2} + \frac{\partial^2\psi}{\partial y^2} + \frac{\partial^2\psi}{\partial z^2}$$

и потенцијал $V(r)$. Таласна функција $\psi(r)$ и енергија E су својствене оператору, што значи да их он само изнова и изнова репродукује, као канал преноса информације који преноси себи адаптирану информацију ни мало је не мењајући при томе. ❑

На питање ми, може ли оператор бити „својствена вредност“ оператора и ако је одговор да како то објашњавам, одговор је позитиван следећим примером, а објашњење ћу изнети након тога.

Пример 4. Временски зависну Шредингерову једначина придодајемо временски независној из прошлог примера. Њена је својствена једначина

$$\partial_t\Psi(x,t) = -\frac{i}{\hbar}\hat{H}\Psi(x,t),$$

где је сада Хамилтонијан (оператор) са фактором $(-i/\hbar)$ својствена вредност, а уз претходну својствену функцију приписан је временски фактор. ❑

Што се математике тиче, оператори су врста вектора, а (једнокомпонентни) вектори су и бројеви. Репрезентација тог формализма на кванту физику, на пример, је да су први од поменутих процеси, други стања, а трећи вредности. То је, пак, само пуко становиште својствено нам, ни мало посебно ако се на васиону гледа са позиција „теорије информације“.

Процеси којима се подвргавају рецимо електрони у кретању под дејством електричног набоја не важе за честице без електричног набоја. Хоћу рећи, процеси су везани за стања (електрона) као и стања за процесе. Што их боље посматрамо видимо дуалност, симетрију између њих, коју алгебра несумњиво доказује. У томе је ствар у бољем схватање те алгебре и њених примена, сматрам.

Додатно, полазећи од стварне неизвесности као структуре информације, а ове као ткива васионе којој припадамо, показује се да постоји онолико димензија времена колико и физичког простора (Dimensions) и, штавише, да су свих шест формално равноправне. Ово последње у смислу да је (са постојећим знањем теоријске физике) могуће, из 6-дим простор-времена васионе, издвојити било које четири димензије и три од њих x, y и z прогласити просторним, а четврту ict временском, и у таквом моделу наћи исте законе физике који важе и за наш.

То би могао бити један једноставан расплет мистерије простора, времена и материје. У поређењу са оваквим информатичким, да оно што могу опажати чини неку „моју реалност“, да још субјеката мени сличних њу „подебљава“ и чини стварност утолико стварнијом, чвршћом, извеснијом што је више учесника могућих доживљаја, тренутно немамо добрих алтернатива. Разматрајући поставке ове теорије информације немојмо олако превидети огромност васионе, стања и процеса, њених „посматрача“ и прелако запасти у „себичност“. Потцењујући значај и количину других могло би нам нестати „магије“ ове теорије.

71.5. Матрица

Степена функција матрице дефинише се са

$$e^{\hat{M}} = \sum_{n=0}^{\infty} \frac{\hat{M}^n}{n!}.$$

Непосредно степеновање матрице није једноставан поступак, осим ако је матрица дијагонална, ако су јој сви елементи нуле сем оних на главној дијагонали. Тада је њен степен једнак матрици степенованих дијагоналних елемената, па је могуће израчунавање и других функција развојем у степене редове.

Квадратна матрица $\hat{M}$ је недефектна (може бити дијагонализована) ако је „слична“ дијагоналној матрици $\hat{A}$. Тада постоји инвертибилна матрица $\hat{B}$ тако да је $\hat{M} = \hat{B}\hat{A}\hat{B}^{-1}$, или еквивалентно томе $\hat{A} = \hat{B}^{-1}\hat{M}\hat{B}$. Затим израчунавамо:

$$\hat{M}^2 = \hat{M}\hat{M} = \left(\hat{B}\hat{A}\hat{B}^{-1}\right)\left(\hat{B}\hat{A}\hat{B}^{-1}\right) = \hat{B}\hat{A}^2\hat{B}^{-1},$$

$$\hat{M}^{n+1} = \hat{M}^n\hat{M} = \left(\hat{B}\hat{A}^n\hat{B}^{-1}\right)\left(\hat{B}\hat{A}\hat{B}^{-1}\right) = \hat{B}\hat{A}^{n+1}\hat{B}^{-1}.$$

Иначе, степен матрице није матрица степенованих коефицијената (осим ако је дијагонална), нити је логаритам матрице једнак матрици логаритама.

Из алгебре знамо да се налажење сличне матрице датој своди на тражење својствених вредности λ и одговарајућих својствених вектора $\vec{x}$, тако да је $\hat{M}\vec{x} = \lambda\vec{x}$. На дијагонали матрице $\hat{A}$ налазе се пронађене својствене вредности, док ће јој сви остали елементи бити нуле.

Пример 5. Налазимо експонент матрице:

$$\hat{M} = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \\ 2 & 2 & 3 \end{pmatrix},$$

$$0 = \det(\hat{M} - \lambda\hat{I}) = \begin{vmatrix} 1-\lambda & 0 & -1 \\ 1 & 2-\lambda & 1 \\ 2 & 2 & 3-\lambda \end{vmatrix}.$$

Након решавања ове кубне једначине налазимо три својствена решења $\lambda = 1, 2, 3$. Даље тражимо њима одговарајуће својствене векторе, редом $\hat{M}\vec{x}_k = \lambda_k \vec{x}_k$, за $\lambda_k \in \{1, 2, 3\}$:

$$\begin{pmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \\ 2 & 2 & 3 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}_k = \lambda_k \cdot \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}_k,$$

$$\begin{pmatrix} 1-\lambda_k & 0 & -1 \\ 1 & 2-\lambda_k & 1 \\ 2 & 2 & 3-\lambda_k \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}_k = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

$$\vec{x}_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \vec{x}_2 = \begin{pmatrix} -2 \\ 1 \\ 2 \end{pmatrix}, \quad \vec{x}_3 = \begin{pmatrix} 1 \\ -1 \\ -2 \end{pmatrix},$$

$$\hat{B} = \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & -1 \\ 0 & 2 & -2 \end{pmatrix},$$

$$\hat{B}^{-1} = \frac{\operatorname{adj}(\hat{B})}{\det(\hat{B})} = \frac{1}{2}\begin{pmatrix} 0 & -2 & 1 \\ -2 & -2 & 0 \\ -2 & -2 & -1 \end{pmatrix}.$$

Непосредним матричним множењем проверавамо да је $\hat{B}\hat{B}^{-1} = \hat{B}^{-1}\hat{B} = \hat{I}$, а то значи да је

$$\hat{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix},$$

$$e^{\hat{M}} = \hat{B}(e^{\hat{A}})\hat{B}^{-1} =$$

$$= \begin{pmatrix} 1 & -2 & 1 \\ -1 & 1 & -1 \\ 0 & 2 & -2 \end{pmatrix} \begin{pmatrix} e^1 & 0 & 0 \\ 0 & e^2 & 0 \\ 0 & 0 & e^3 \end{pmatrix} \begin{pmatrix} 0 & -1 & 1/2 \\ -1 & -1 & 0 \\ -1 & -1 & -1/2 \end{pmatrix}.$$

Исти је поступак за израчунавање логаритма матрице, када уместо e^{λ} завршавамо са $\log\lambda$. ❑

Логаритам матрице $\hat{M}$ може бити комплексна матрица чак и када је $\hat{M}$ реална, јер матрица са реалним и позитивним коефицијентим може имати негативне или чак комплексне својствене вредности (на пример матрица ротације). Отуда долази и нејединственост логаритма матрице, која је последица нејединствености (периодичности) логаритма комплексног броја.

72. Слични оператори

Линеарни оператор f обично дефинише адитивност $(\forall x_1, x_2 \in X)\ f(x_1 + x_2) = f(x_1) + f(x_2)$ и хомогеност $f(\alpha x) = \alpha f(x)$, што важи за сваку промењљиву (x) из домена (X) и сваки тзв. скалар $(\alpha \in \Phi)$. Адитивност и хомогеност дефинишу линеарност

$$(\forall x, y \in X)(\forall \alpha, \beta \in \Phi)\ f(\alpha x + \beta y) = \alpha f(x) + \beta f(y).$$

Множење константом и извод линеани су оператори, јер:

$$C \cdot (\alpha f + \beta g) = \alpha C \cdot f + \beta C \cdot g \quad \text{и} \quad (\alpha f + \beta g)' = \alpha f' + \beta g',$$

као и интеграл. Зато је и Хамилтонијан

$$\hat{H} = -\frac{\hbar^2}{2m}\nabla^2 + V(\vec{r})$$

линеарни оператор.

Захваљујући дефиниционим особинама, домен линеарне функције може се испарцелисати на низ сабирака, на различите начине. Кажемо да низ $e_1, e_2, \dots, e_n$ разапиње такав један простор X, да је база, ако се свако $x \in X$ може писати као линеарна комбинација елемената низа, дакле ако је

$$x = \sum_{k=1}^{n} \xi_k e_k, \quad \xi_1, \xi_2, \dots, \xi_n \in \Phi.$$

Тада и слике датог низа $f_k = L(e_k)$, редом за $k = 1, 2, \dots, n$, било којег линеарног оператора истог домена, $L: X \to Y$, разапињу простор Y слика, односно простор кодомена оператора. Наиме:

$$Lx = \sum_{k=1}^{n} \xi_k L e_k = \sum_{k=1}^{n} \xi_k f_k.$$

Поред тога видимо да је деловање оператора потпуно одређено деловањем на базу простора, да су два оператора једнака ако на њу једнако делују.

Јединствено одређење линеарног оператора A пресликавањем база простора изражава релација

$$Ae_j = \sum_{i=1}^{m} \alpha_{ij} f_i, \quad j = 1, 2, \dots, n$$

која елементу $A \in (X^n \to Y^m)$ придружује једну и само једну матрицу $\hat{A} = (\alpha_{ij})_{n \times m}$. Ова релација се може користити као дефиниција матрица, за доказивање особина матрица а затим и линеарних оператора. На пример, множење матрица (композиција оператора) асоцијативно је али није увек комутативно.

Пример 1. Дате су две матрице другог реда:

$$\hat{A} = \begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix},$$

$$\hat{A}\hat{B} = \begin{pmatrix} -4 & -11 \\ -8 & -19 \end{pmatrix}, \quad \hat{B}\hat{A} = \begin{pmatrix} -7 & 4 \\ 31 & -16 \end{pmatrix}.$$

Како је $\hat{A}B \neq \hat{B}\hat{A}$, оне нису комутативне. ❑

Процеси, њихови оператори и матрице, нису комутативни када је битан редослед радњи. Није нам свеједно решавамо ли неки тест па рецимо једемо бунике, или урадимо обрнуто. Ако први процес утиче на следећег, онда се може десити некомутативност композиције (производа) њих два. Боље речено, када су токови независни један од другога они су тада комутативни. Као бацање новчића, где следећи исход не зависи од претходних, за комутативне процесе можемо рећи да „не памте“ своја претходна стања, или да су константни.

Тако је комутативна матрица сама са собом, јер је константна, али и процес који је реверзибилан комутира са својим инверзним, јер се у композицији њих два стално враћају на исти почетак, или крај. Познате су нам многе теореме које ово доказују, па не шкоди да их разумемо и интуитивно.

72.1. Реверзибилност

Нека је производ две матрице јединична матрица, или замислимо композицију процеса и њему инверзног процеса, линеарног оператора и њему инверзног. Такве ће једнаке векторе преводити само у једнаке векторе; једнаке копије даваће само једнаки оригинали. Наиме, реверзибилност претпоставке значи управо то, да из једнакости слика произилази једнакост ликова. О томе је реч у следећим лемама (малим теоремама), у коначно димензионалним векторским просторима.

Лема 1. Ако је $\hat{A}\hat{B} = \hat{I}$, онда из $\hat{B}\vec{x} = \hat{B}\vec{y}$ следи $\vec{x} = \vec{y}$.

Доказ: Следи из низа импликација:

$$\hat{B}\vec{x} = \hat{B}\vec{y} \;\Rightarrow\; \hat{B}(\vec{x} - \vec{y}) = 0 \;\Rightarrow\; \hat{A}\hat{B}(\vec{x} - \vec{y}) = 0 \;\Rightarrow\; \hat{I}(\vec{x} - \vec{y}) = 0 \;\Rightarrow\; \vec{x} - \vec{y} = 0.$$

Дакле $\vec{x} = \vec{y}$ и тврђење је доказано. ∎

Дакле, пресликавање $\hat{B}$ је бијективно (из копија следе јединствени оригинали), па је одговарајући систем линеарних једначина $\hat{B}\vec{x} = \vec{y}$ регуларан (има јединствено решење), ранг матрице $\hat{B}$ једнак је броју непознатих, а њена детерминанта није нула.

Пример 2. Нека је ова друга матрица једнака инверзној првој из 1. примера:

$$\hat{A} = \begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix}, \quad \hat{B} = \hat{A}^{-1} = \frac{\operatorname{adj} \hat{A}}{\det \hat{A}} = \frac{1}{-6}\begin{pmatrix} 4 & -2 \\ 7 & -5 \end{pmatrix}.$$

Тада из $\hat{B}\vec{x} = \hat{B}\vec{y}$ следи:

$$\begin{pmatrix} 4 & -2 \\ 7 & -5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 & -2 \\ 7 & -5 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

$$\begin{cases} 4x_1 - 2x_2 = 4y_1 - 2y_2 \\ 7x_1 - 5x_2 = 7y_1 - 5y_2 \end{cases}$$

$$\begin{cases} 4(x_1 - y_1) - 2(x_2 - y_2) = 0 \\ 7(x_1 - y_1) - 5(x_2 - y_2) = 0 \end{cases}$$

$$x_1 - y_1 = 0, \quad x_2 - y_2 = 0,$$

дакле, $\vec{x} = \vec{y}$. ❑

Лема 2. Ако је $\hat{A}\hat{B} = \hat{I}$, онда из линеарне независности вектора $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ приизилази линеарна независност вектора $\vec{v}_1 = \hat{B}\vec{u}_1, \vec{v}_2 = \hat{B}\vec{u}_2, \dots, \vec{v}_n = \hat{B}\vec{u}_n$ (ни један од линеарно независних вектора се не може представити као линеарна комбинација осталих).

Доказ: Ако су вектори $\vec{u}_k$ линеарно независни, онда из $\sum_{k=1}^{n} \alpha_k \vec{u}_k = \vec{0}$ следи да је сваки од скалара α_k једнак нули. Дакле, биће $\sum_{k=1}^{n} \alpha_k \vec{u}_k \neq \vec{0}$, ако бар један скалар није нула. На основу претходне леме је $\vec{0} = \hat{B}\vec{0} \neq \hat{B} \sum_{k=1}^{n} \alpha_k \vec{u}_k = \sum_{k=1}^{n} \alpha_k \hat{B}\vec{u}_k = \sum_{k=1}^{n} \alpha_k \vec{v}_k$, а то је линеарна независност вектора $\vec{v}_k$ такође. ■

Пример 3. Матрице из 1. примера имају следеће својствене једнакости:

$$\hat{A}: \quad \begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix} \begin{pmatrix} 2 \\ 7 \end{pmatrix} = 2 \begin{pmatrix} 2 \\ 7 \end{pmatrix}, \quad \begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -3 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\hat{B}: \quad \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} = -\begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \end{pmatrix} = -2 \begin{pmatrix} 1 \\ -2 \end{pmatrix}$$

што је лако проверити. Својствени вектори прве независни су, а такође и друге. ❑

У условима претходне леме, када су вектори $\vec{u}_k$ базни простора X, а њихове слике $\vec{v}_k = \hat{B}\vec{u}_k$ базни простора Y, онда је произвољан вектор $\vec{y} \in Y$ облика

$$\vec{y} = \sum_{k=1}^{n} \alpha_k \vec{v}_k = \sum_{k=1}^{n} \alpha_k \hat{B}\vec{u}_k = \hat{B} \sum_{k=1}^{n} \alpha_k \vec{u}_k = \hat{B}\vec{x}.$$

Према томе, за сваки вектор $\vec{y} \in Y$ постоји вектор $\vec{x} \in X$ такав да је $\vec{y} = \hat{B}\vec{x}$. Отуда следећа лема.

Теорема 3. Ако је $\hat{A}\hat{B} = \hat{I}$, онда је $\hat{B}\hat{A} = \hat{I}$.

Доказ: Из датог услова следе претходне леме и додатно:

$$(\hat{B}\hat{A} - \hat{I})\vec{y} = (\hat{B}\hat{A} - \hat{I})\hat{B}\vec{x} = (\hat{B}\hat{A}\hat{B} - \hat{B})\vec{x} = [\hat{B}(\hat{A}\hat{B} - \hat{I})]\vec{x} = (\hat{B}\vec{0})\vec{x} = \vec{0}\vec{x} = \vec{0}.$$

Отуда $\hat{B}\hat{A} = \hat{I}$, па је теорема доказана. ■

Теорема утврђује да за инверзне матрице, па и за линеарне операторе, важи закон комутације. То је оно што смо претходно разумели интуитивно, а даље погледајмо како се исто може доказати и самим свођењем на контрадикцију. Ако претпоставимо да је $\hat{A}\hat{B} = \hat{I}$ и $\vec{x}$ неки вектор којег $\hat{B}\hat{A}$ неће пресликати у самог себе, онда је:

$$\hat{B}\vec{x} \neq \hat{B}\hat{A}(\hat{B}\vec{x}) = \hat{B}(\hat{A}\hat{B}\vec{x}) = \hat{B}(\hat{I}\vec{x}) = \hat{B}\vec{x},$$

а то је контрадикција. Претпоставка је нетачна, не постоји такав вектор, тј. $\hat{B}\hat{A} = \hat{I}$ такође.

Резимирајмо. Узајамно инверзни линеарни оператори, процеси које представљају као и припадне матрице, комутативни су. Они представљају бијекције (обострано једнозначна пресликавања), па су и векторски простори њима пресликани једнако димензионални. Колоне (врсте) такве матрице чине линеарно независни вектори.

72.2. Дијагонализација

Дијагонална матрица $\hat{D}$ је квадратна која само на главној дијагонали има ненулте елементе. Она је комутативна са сваком матрицом $\hat{A}$ истог реда. То је очигледно, ако имамо у виду примере попут:

$$\begin{cases} \hat{A}\hat{D} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} = \begin{pmatrix} a & 2b & 3c \\ 4a & 5b & 6c \\ 7a & 8b & 9c \end{pmatrix}, \\ \hat{D}\hat{A} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} a & 2b & 3c \\ 4a & 5b & 6c \\ 7a & 8b & 9c \end{pmatrix}. \end{cases}$$

Поготово су две дијагоналне матрице истог реда увек комутативне.

Дијагонална матрица са свим ненултим дијагоналним елементима има инверзну (регуларна је):

$$\hat{D}\hat{D}^{-1} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}\begin{pmatrix} a^{-1} & 0 & 0 \\ 0 & b^{-1} & 0 \\ 0 & 0 & c^{-1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

када је $abc \neq 0$.

Својствене вредности дијагоналних матрица управо су њихови дијагонални елементи. То је такође очигледно, на основу примера:

$$\hat{D}\vec{x} = \lambda\vec{x} \quad \Rightarrow \quad (\hat{D} - \lambda\hat{I})\vec{x} = 0 \quad \Rightarrow$$

$$\begin{vmatrix} a-\lambda & 0 & 0 \\ 0 & b-\lambda & 0 \\ 0 & 0 & c-\lambda \end{vmatrix} = 0 \quad \Rightarrow$$

$$(a-\lambda)(b-\lambda)(c-\lambda) = 0,$$

а отуда, својствене вредности су $\lambda_1 = a$, $\lambda_2 = b$ и $\lambda_3 = c$. Одговарајући својствени вектори следе из једначина $\hat{D}\vec{x}_k = \lambda_k\vec{x}_k$, за индексе $k = 1, 2, 3$. Први је:

$$\hat{D}\vec{x}_1 = \lambda_1\vec{x}_1 \quad \Rightarrow \quad \hat{D}\vec{x}_1 = a\vec{x}_1 \quad \Rightarrow$$

$$\begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}_1 = a\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix}_1 \quad \Rightarrow \quad \begin{pmatrix} a\xi_1 \\ b\xi_2 \\ c\xi_3 \end{pmatrix}_1 = \begin{pmatrix} a\xi_1 \\ a\xi_2 \\ a\xi_3 \end{pmatrix}_1 \quad \Rightarrow \quad \vec{x}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

ако су a, b и c три различита броја. Слично налазимо:

$$\vec{x}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{x}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

тако да је $\hat{D}\vec{x}_2 = b\vec{x}_2$ и $\hat{D}\vec{x}_3 = c\vec{x}_3$. Приметимо да матрица колона својствених вектори дијагоналне матрице мора бити или јединична, или је нека пермутација колона јединичне.

Мало је сложенија ситуација са произвољном инвертибилном квадратном матрицом $\hat{A}$ коју треба добити множењем са дијагоналном.

Лема 4. Нека је $\hat{A}$ квадратна матрица n-тог реда са својственим вредностима λ_1, λ_2, ..., λ_n и њима припадним својственим векторима $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$, тако да је $\hat{A}\vec{v}_k = \lambda_k \vec{v}_k$ за $k = 1, 2, \dots, n$. Нека је $\hat{C}$ матрица чије колоне су ти својствени вектори, а $\hat{D}$ је дијагонална матрица чију главну дијагоналу чине редом наведене својствене вредности. Тада и само тада је $\hat{A}\hat{C} = \hat{C}\hat{D}$.

Доказ: Под датим условима:

$$\hat{A}\hat{C} = \hat{A}(\vec{v}_1 \quad \vec{v}_2 \quad \dots \quad \vec{v}_n) = \left(\hat{A}\vec{v}_1 \quad \hat{A}\vec{v}_2 \quad \dots \quad \hat{A}\vec{v}_n\right),$$

$$\hat{C}\hat{D} = (\vec{v}_1 \quad \vec{v}_2 \quad \dots \quad \vec{v}_n)\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} = (\lambda_1\vec{v}_1 \quad \lambda_2\vec{v}_2 \quad \dots \quad \lambda_n\vec{v}_n),$$

а отуда $\hat{A}\hat{C} = \hat{C}\hat{D}$. Тиме је став доказан. ■

Приметимо да ова лема доказује исправност поступка у недавном примеру [71.5. Матрица].

Теорема 5. Матрица n-тог реда може се дијагонализирати ако и само ако има n линеарно независних својствених вектора.

Доказ: Нека је то квадратна матрица $\hat{A}$, нека она има n линеарно независних својствених вектора. Тада је регуларна матрица $\hat{C}$, претходне леме, чије су колоне ти вектори, њен је ранг n.

Обрнуто, ако се $\hat{A}$ може дијагонализовати, онда, по дефиницији, постоји регуларна матрица $\hat{C}$ таква да је $\hat{D} = \hat{C}^{-1}\hat{A}\hat{C}$ дијагонална. Претходна лема тада каже да ће колоне $\hat{C}$ бити својствени вектори $\hat{A}$. Како је $\hat{C}$ регуларна, или што је исто инвертибилна, то њених n колона морају бити линеарно независни вектори. Према томе, $\hat{A}$ има n линеарно независних својствених вектора. ■

72.3. Комутативност

Видели смо да су независни процеси комутативни. То су они који се узајамно не узурпирају, затим таква су два корака истог константног процеса, кажемо оног који не памти, или помало неспретно названог „акутног" процеса који је једнократан и не акумулира се, за разлику од „хроничног" чије је дејство дуже и погоршавајуће. Следећи став открива такве међу матрицама.

Теорема 6. Две матрице које се истовремено могу дијагонализовати су увек комутативне.

Доказ: Нека су $\hat{A}$ и $\hat{B}$ две такве квадратне матрице реда n са истим низом својствених вектора $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_n$. Како се $\hat{A}$ и $\hat{B}$ могу симултано дијагонализовати, постоји база вектора заједничка за обе. Означимо припадне својствене вредности матрице $\hat{A}$ са $\alpha_1, \alpha_2, \dots, \alpha_n$, а матрице $\hat{B}$ са β_1, β_2, ..., β_n, тј. нека је $\hat{A}\vec{v}_k = \alpha_k\vec{v}_k$ и $\hat{B}\vec{v}_k = \beta_k\vec{v}_k$ редом за $k = 1, 2, \dots, n$.

Тада је матрица $\hat{C}$, чије колоне су вектори $\vec{v}_1$, $\vec{v}_2$, ..., $\vec{v}_n$, таква да је $\hat{A} = \hat{C}\hat{D}_\alpha\hat{C}^{-1}$ и $\hat{B} = \hat{C}\hat{D}_\beta\hat{C}^{-1}$, где су $\hat{D}_\alpha$ и $\hat{D}_\beta$ дијагоналне матрице својствених вредности α_k и β_k матрица $\hat{A}$ и $\hat{B}$ дуж дијагонала. Да $\hat{D}_\alpha$ и $\hat{D}_\beta$ комутирају јасно је (72.2), а затим налазимо:

$$\hat{A}\hat{B} = \left(\hat{C}\hat{D}_\alpha\hat{C}^{-1}\right)\left(\hat{C}\hat{D}_\beta\hat{C}^{-1}\right) = \hat{C}\hat{D}_\alpha(\hat{C}^{-1}\hat{C})\hat{D}_\beta\hat{C}^{-1} = \hat{C}\hat{D}_\alpha\hat{D}_\beta\hat{C}^{-1} =$$

$$= \hat{C}\hat{D}_\beta\hat{D}_\alpha\hat{C}^{-1} = \hat{C}\hat{D}_\beta(\hat{C}^{-1}\hat{C})\hat{D}_\alpha\hat{C}^{-1} = \left(\hat{C}\hat{D}_\beta\hat{C}^{-1}\right)\left(\hat{C}\hat{D}_\alpha\hat{C}^{-1}\right) = \hat{B}\hat{A}.$$

Тиме је теорема доказана. ■

Када сигнал жмигањем на раскрсници у саобраћају не би имао значаја, онда би било свеједно да ли возач прво укључи жмигавац па скрене у бочну улицу, или прво скрене у бочну улицу па најави скретање. Тада би жмигање и скретање толико много припадали истом „свету“ (простору вектора) да није потребно „подебљавање“ (понављање, наглашавање) тада заправо „једне те исте“ радње. Међутим, знамо да постоје саобраћајне несреће због непоштовања сигнализације пре скретања, а то онда значи да те две радње припадају делимично различитим „световима“.

Оператори потпуно различитих векторских простора (наводних светова) опет су комутативни. Онај пресек таквих, који чини да та два простора интерагују (узурпирају један другог), учиниће процесе некомутативним.

Пример 4. Систем линеарних једначина

$$\begin{cases} a_{11}x_1 + \dots + a_{1n}x_n = y_1 \\ \dots \\ a_{n1}x_1 + \dots + a_{nn}y_n = y_n \end{cases}$$

интерпретација је линеарног пресликавања $A: \vec{x} \to \vec{y}$, као и матричне једначине ($\hat{A}\vec{x} = \vec{y}$). Такав систем има јединствено решење ако и само ако његова главна детерминанта, или детерминанта матрице, није нула ($\det \hat{A} \neq 0$). Матрица система је регуларна, значи инвертибилна и има колоне (врсте) линеарно независне векторе.

Када две регуларне матрице истог реда, са по n вектора колона и по n компоненти сваког од тих вектора, имају заједнички сопствени вектор – оне су комутативне. Наиме, из $\hat{A}\vec{v} = \alpha\vec{v}$, $\hat{B}\vec{v} = \beta\vec{v}$ и регуларности матрица $\hat{A}$ и $\hat{B}$, биће:

$$\hat{A}\hat{B}\vec{v} = \hat{A}(\hat{B}\vec{v}) = \hat{A}(\beta\vec{v}) = \beta(\hat{A}\vec{v}) = \beta\alpha\vec{v},$$

$$\hat{B}\hat{A}\vec{v} = \hat{B}(\hat{A}\vec{v}) = \hat{B}(\alpha\vec{v}) = \alpha(\hat{B}\vec{v}) = \alpha\beta\vec{v},$$

а затим и комутативност ових матрица. ❑

Један од најпростијих примера нерегуларних некомутативних матрица је:

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Иначе је комутативно множење произвољне матрице (истог реда) са јединичном матрицом, нула матрицом и дијагоналном матрицом. Производ две матрице ротација комутативан је.

Славни пример некомутативности у квантној механици су Хајзенбергове релације неодређености. Када се фотоном (светлошћу) гађа електрон ради одређивања му положаја, преноси му се импулс. Што је мања таласна дужина фотона то му је већи импулс, тада је тачније одређен положај мете, а нетачнији увид у импулс. Положај и импулс електрона, а тако и енергија и време, у просторима су који се тако делимично преклапају да су њихови процеси некомутативни.

Некомутативност неких линеарних оператора (матрица) неотклоњива је, принципијелна је ствар алгебре која доводи до принципа неодређености. Отуда је и неодређеност која је постулирана у (мојој) „теорији информације“ принципијелна ствар која наговештава да је та теорија на добром путу.

Неодређеност је релативна. Зато комуницирамо (интерагујемо) јер немамо све информације које нам требају. Ловац спрема клопке за дивљач која не зна шта јој се може десити. Неодређеност је слојевита. Она напредује као што савремена цивилизација „осваја“ науку, корак по корак, упорно али никада не стижући до коначног циља. Нека неодређеност увек постоји изван ма како великог скупа знања, каже нам Геделова теорема немогућности.

Ево на крају и једног занимљивијег од алгебарских доказа принципа неодређености[6]. Очекивање, или средња вредност случајне променљиве, овде обзервабле A (физички мерљиве величине коју може представљати и линеарни оператор, матрица или реална својствена вредност) је $\mu_A = \langle A\rangle$. Варијанса, или средња вредност одступања променљиве од средње вредности је $\Delta A = \langle A - \mu_A\rangle$. Исто важи за обзерваблу B.

Пример 5. Корени варијансе, дисперзије варијабли $\sigma_A = \Delta A$ и $\sigma_B = \Delta B$, њихове су неодређености. Додатно, оне могу деловати на таласну функцију ψ као оператори. Тако је:

$$\|(\Delta A + i\lambda\Delta B)\psi\|^2 \ge 0,$$

$$\langle\psi, (\Delta A - i\lambda\Delta B)(\Delta A + i\lambda\Delta B)\psi\rangle \ge 0,$$

$$\langle\psi, \Big((\Delta A)^2 + \lambda^2(\Delta B)^2 + i\lambda(\Delta A\Delta B - \Delta B\Delta A)\Big)\psi\rangle \ge 0,$$

$$(\Delta A)^2 + \lambda^2(\Delta B)^2 + i\lambda[\Delta A, \Delta B] \ge 0.$$

Овде је комутатор:

$$[\Delta A, \Delta B] = \Delta A\Delta B - \Delta B\Delta A =$$

$$= \langle A - \mu_A\rangle\langle B - \mu_B\rangle - \langle B - \mu_B\rangle\langle A - \mu_A\rangle$$

$$= (AB - A\mu_B - B\mu_A + \mu_A\mu_B) - (BA - B\mu_A - A\mu_B + \mu_A\mu_B)$$

$$= AB - BA = [A, B].$$

Сменимо ово горе и добијамо квадратну једначину свугде ненегативну, по непознатој λ. Њена дискриминанта зато је ненегативна и отуда

$$\Delta A\ \Delta B \ge \left|\frac{i}{2}[A, B]\right|.$$

Ово је стандардна Хајзенбергова релација неодређености. У 5. примеру наслова „70.1. Вектори“ је израчунат комутатор импулса и положаја $[\hat{p}, \hat{x}] = -i\hbar$, па је $\Delta\hat{p}\ \Delta\hat{x} \ge \hbar/2$. ❑

[6] C. Cohen-Tannoudji, B. Diu, and F. Laloë, Quantum Mechanics (John Wiley and Sons, New York, 1977), Vol. 1, pp. 286-287.

72.4. Комуникација

Производ неодређености две обзервабле реда је величине комутатора. Као што смо видели, они говоре и о површинама, о оне о информацијама [30. Комутатори]. Те ће се идеје уз Хајзенбергове неодређености, које онемогућавају крајње детерминисање интеракција, показати згодне за даље ширење појма информације, као што је најављено у књизи Информација Перцепције (2016).

Неодређеност положаја може се посматрати и као густина вероватноће налажења честице-таласа на датом месту када она појашњава Комптонов ефекат (Простор-Време, 2017). Препознајемо га и као промену путање електрона на мање вероватну, не без дејства силе (судара са фотоном). Томе додајмо да производ неодређености импулса и положаја може бити и мера комуникације честице на путањи. Укратко, комутатори су неодређености који информацији дају на значају.

Елементарне опције елементарне честице такве су јер их можемо налазити у мноштву код већих, сложенијих система. То је претпоставка, наравно, а отуда информација перцепције:

$$S = \Delta A_1 \Delta B_1 + \Delta A_2 \Delta B_2 + \cdots + \Delta A_n \Delta B_n$$

предмета који комуницирају на n начина. Ови сабирци су неодређености, односно дејства, реда величина комутатора обзервабли A_k и B_k, редом за $k = 1, 2, \dots, n$.

Информација је нарочита мера реализоване „количине неизвесности“. Теорија каже да тај износ не може бити нулти нити за једну реалну честицу, у произвољном њеном „тренутку“ (интервалу времена унутар којег је није могуће тачно детермнисати). Она „честица“ која не комуницира, за нас је псеудо-реална, или непостојећа, јер способност преноса информације, било посредно или непосредно, између два пријемника потврдаје њихове узајамне реалности. То је једна од првих, давних ставки са којом са започео „теорију информације“.

Напомињем још једном, посебност псеудо-реалности, или паралелне реалности, или имагинарне реалности (Bypass) је способност реалне честице да „изабере“ одлазак тамо као и повратак у нашу стварност. У тој важе закони одговарајући нашим, а сам процес „заобилажења“, већ сам исувише помињао (нпр. Lateral). За разлику од честица псеудо-реалности, оне „непостојеће“ честице не би имале нити ту мајушну могућност појављивања у нашој реалности, или шетања из и у њу.

Већа таква „заобилажења“, или простије индивидуалне одласке, односно доласке из оне у нашу реалност, ограничава закон великих бројева теорије вероватноће.

73. Матрица расподела

Матрица чији су ретци (колоне) расподеле вероватноћа назива се стохастичка. Ако су то и ретци и колоне, ненегативни бројеви јединичног збира, она се назива дупла стохастичка матрица. Канали преноса информација обично су такве.

Пример 1. Дата је квадратна дупла стохастичка матрица другог реда

$$\hat{K} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$$

где је $a + b = 1$ и $a, b \geq 0$. Квадрат ове матрице

$$\hat{K}^2 = \begin{pmatrix} a^2 + b^2 & 2ab \\ 2ab & a^2 + b^2 \end{pmatrix}$$

опет је дупла стохастичка матрица, јер $a^2 + b^2 + 2ab = (a + b)^2 = 1$. Тада је

$$\frac{1}{2} < a^2 + b^2 < a < 1,$$

$$\frac{1}{2} < a^2 + (1 - a)^2 < a < 1,$$

што је тачно, јер решења квадратне неједначине $2a^2 - 3a + 1 < 0$ управо су бројеви $a \in (1/2, 1)$ датог интервала. ❑

Пример 2. Дупле стохастичке матрице ($a + b = 1$ и $a, b \geq 0$, $p + q = 1$ и $p, q \geq 0$) комутативне су:

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix}\begin{pmatrix} p & q \\ q & p \end{pmatrix} = \begin{pmatrix} ap + bq & aq + bp \\ bp + aq & pq + ap \end{pmatrix},$$

$$\begin{pmatrix} p & q \\ q & p \end{pmatrix}\begin{pmatrix} a & b \\ b & a \end{pmatrix} = \begin{pmatrix} pa + qb & pb + qa \\ qa + pb & qb + pa \end{pmatrix},$$

а њихови производи:

$$(ap + bq) + (aq + bp) = a(p + q) + b(q + p) = a + b = 1$$

увек су неке дупле стохастичке матрице. ❑

Први пример показује да дупла стохастичка матрица множењем (надовезивањем у ланац) постаје све ближа „црној кутији“, међусобно једнаких вероватноћа својих коефицијената. Настављањем у дуги низ канала она даје све дужи ланац, такође канала:

$$\hat{K}, \hat{K}^2, \hat{K}^4, \dots, \hat{K}^{2n} \to \hat{K}^{\infty}, \quad \text{када } n \to \infty.$$

Гранични, далеки канал ($\hat{K}^{\infty}$) има уједначене коефицијенте, само јединицу за сопствену вредност и само један сопствени (својствени, карактеристични) вектор:

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}.$$

То смо доказивали различитим теоремама раније, а подсећамо се ради лакшег разумевања тих случајева у недавним резултатима о сличним операторима.

Црна кутија је канал који гуши и не саопштава информацију, па кажемо да је „не преноси“,

$$\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix},$$

када је $p + q = 1$. Системи које оне представљају међусобно не комуницирају, нити комуницирају са двоструким стохастичким каналима.

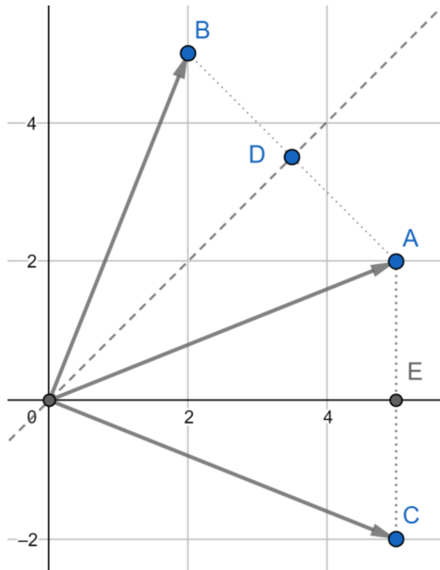
Пример 3. Израчунавамо комутатор „црне кутије“ и произвољне матрице другог реда:

$$[\hat{K}^\infty, \hat{M}] = \hat{K}^\infty \hat{M} - \hat{M}\hat{K}^\infty =$$

$$= \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$$

$$= \frac{1}{2}\begin{pmatrix} c-b & d-a \\ a-d & b-c \end{pmatrix}.$$

Резултат је нула матрица, односно комутативност ако и само ако $c - b = d - a = 0$. Дакле, да би црна кутија комуницирала са матрицом $\hat{M}$ (другог реда) потребно је и довољно да та бар на једној од две дијагонале нема једнаке елементе. ❑



Матрице ротације нису комутативне са $\hat{K}^\infty$. Занимљиво је да комутатор горњег примера добијамо множењима прве Паулијеве матрице [70.1. Вектори].

Пример 4. Израчунавамо комутатор Паулијеве и било које матрице другог реда:

$$[\hat{P}_1, \hat{M}] = \hat{P}_1\hat{M} - \hat{M}\hat{P}_2 =$$

$$= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} - \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} c & d \\ a & b \end{pmatrix} - \begin{pmatrix} b & a \\ d & c \end{pmatrix} = \begin{pmatrix} c-b & d-a \\ a-d & b-c \end{pmatrix},$$

а половина овога је комутатор 3. примера. ❑

Паулијева прва матрица интерпретација је рефлексије вектора око симетрале првог квадранта ($\hat{P}_1: \overrightarrow{OA} \to \overrightarrow{OB}$), а трећа око x осе ($\hat{P}_3: \overrightarrow{OA} \to \overrightarrow{OC}$). То демонстрира слика лево:

$$\hat{P}_1\vec{v} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}, \quad \hat{P}_3\vec{v} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}.$$

Укратко, процеси су врста „стања“ са којима друга стања могу комуницирати. То је кратак и нагли заокрет са интеракције на информацију о којем вреди расправљати.

***

Питање: Можете ли ми појаснити о каквим се то „комуницирањима“ ради...?

Одговор: Матрице представљају линеарне операторе, ови процесе, развоје или промене стања рецимо квантне механике (тамо је таква метода скоро једина и изванредно је тачна), али и иначе. Разлику променом редоследа операција називам „комутатором“, а таква за импулс и положај, или енергију и време, представља Хајзенбергове релације неодређености.

Физичка димензија поменутог комутатора је физичко дејство, он сам је квант дејства, минимална интеракција физичког света, али у (мојој) теорији информације исти је и еквивалент минималној „комуникацији“. Оно што не комуницира не интерагује (са нашом реалношћу), а интерагују стања са стањима (честице и честице), па и процеси са стањима и са процесима.

Електрон нерадо емитује своју неодређеност у облику информације, али се нађе у изнудици када га погоди фотон и бива принуђен да комуницира. Честица рецимо ода део свог положаја, у замену за део неодређености тада импулса. Мерење је интеракција предмета посматрања са апаратуром, а сада само додајемо да се ту ради и о еквивалентном преносу информација тамо и онамо.

Расправљано је (доказивано) такође да „црна кутија“, неми канал преноса информације, али која заправо гуши и не саопштава примљено, па у том смислу дезинформације кажемо не прослеђује информацију, комуницира на начин Паулијевих матрица. Оне су само детаљ ширих комуникација са ротацијама. Што се тиче ових других, знамо да се свака количина за коју важи закон одржања (материје, енергије, информације, у геометрији изометрије), може свести на, представити самим ротацијама.

Утолико поистовећивање комутатора са комуникацијом постаје значајније. Док некомутативност процеса говори о некој „количини неизвесности“ (информацији) коју они међусобно размењују, дакле о интеракцији међу процесима, дотле комутативност говори о одсуству интеракције, па и о томе да нема комуникације.

Питање: Канал који не преноси информацију еквивалент је Паулијевим матирцама?

Одговор: Тако некако. Тачније „црна кутија“, или канал који преноси превише информације, који зато дезинформише гушећи је, комуницира са бинарним процесима (матрице другог реда) онако како би то радила прва Паулијева матрица…

***

Питање: Да ли сте приметили да је сам комутатор антикомутативан и да је формално фермион?

Одговор: Да, таква је разлика парова производа типична за фермионе[7], али ту сам причу оставио за касније. Оно са чиме се комуницира су честице за које важи Паулијев принципу искључења, што подразумева да дати систем нема две идентичне те врсте, два фермиона, који би се у исто време нашли у потпуно истом стању. То је у складу са становиштем да је неизвесност бит информације.

Са друге стране, то је у складу и са начелном штедљивошћу емисије информације. Наиме, спонтан раст ентропије супстанце прати такође спонтан пад информације датог система. Тежња за мањим титрањем молекула гаса (који би да се хлади) уједно је и тежња за мање њихове комуникације. А

[77] Fermions, Half-integer spin particles; https://www.academia.edu/9994343

затим долази мало већа (просечна) вероватноћа преласка фермиона у бозоне, него обрнуто, што је наставак ове идеје о којој сам већ писао и који бих за сада прескочио (Big Bang).

***

Питање: Ваши прилози теорији информације били би и даље оригинални, али популарнији и интересантнији многима, када бисте их преносили и на друштвене појаве, зар не?

Одговор: Да, слажем се и то понекад урадим, додуше на мој апстрактан начин. Погледајте „Приче о Информацији“. Друштво је нешто између „живог“ и „неживог“ бића, скалирано информацијама (количином опција, могућностима бирања, способностима деловања), а тек затим нешто друго, са становишта моје теорије. Живо биће то је што је јер има вишак информације у односу на неживо и зато већу слободу деловања.

За сву нама познату мртву твар физике важи принцип најмањег дејства; све трајекторије кретања честица, од најмањих честица-таласа до небеских тела, могу се извести из тог принципа, из Ојлер-Лагранжових једначина кретања. Мртва се твар уопште не одупире, па ипак има неке своје изборе видљиве на микро нивоу квантне механике, који постепено, под законима теорије вероватноће и, пре свега закона великих бројева, постају извесности макро света. Користећи нама недавно знане ефектре теорије хаоса (лептирових крила), или тако некако, жива бића успевају настати, сачувати и овладати том врстом слобода микро света у свом макро свету.

Зато су цивилизације животније што поседују више слобода својствених микро свету. За виталност кључно је имати слободе. Али тек трошећи их и зато трошећи се. Моја „теорија информације“ (она је у настанку) дефинише развојност уопште живих бића на јединствен и нов начин. Мера слободе, на начин специфичне информације и њених релативних промена, разликује периоде „младости“, затим „зрелости“, потом „старости“, према расту, стагнацији и опадању тих мера.

Међутим, фокусирана је слобода (количина опција, информација, потрошена неизвесност) та која „разгрће“ ствари око и испред себе постижући развојност, ефикасност и сигурност. Неизвесност је сила (у апстрактном, али и реалном физичком смислу) која се тада троши, а од које би сва природа да побегне (начелна јој је штедљивост емисије информације). То је она дубока, нама несхватљива тежња која покреће наше жеље ка реду, откривању природних закона и додавању наших њима, а системи веће „специфичне информације“ – да емисије информације (неизвесности) буде мање.

Принципијелном минимализму информације опире се закон одржања, али живот ипак успева да полако и упорно надвладава вишкове виталности, да је топи у сигурност, а ову даље у почетно и најприродније могуће стање тоталног неодупирања, у мртву твар. Као што видите, ова теорија учи да је неред резервоар живота, а ред и сигурност да су знаци његове потрошености. Можда таквој управо добро стоји непопуларност и скривање у домену апстракција?

74. Сепарабилни процес

На аеродрому се налази $n = 0, 1, 2, \ldots$ путника. Вероватноћа да ће у аеродром ући $(n+1)$-и је a_n, док је вероватноћа да ће један изаћи b_n. Претпоставимо да смо посматрањем нашли статистичке вероватноће $a_0, a_1, a_2, \ldots$ и $b_0, b_1, b_2, \ldots$ $(b_0 = 0)$. Постављамо питање колика је вероватноћа p_n да ће се на аеродрому наћи тачно n путника. Сепарабилност овде значи да је увек могуће установити ко је од путника први ушао, односно изашао, уколико се на вратима истовремено појаве два.

То је модел који се може применити на процесе рођења и смрти, тачке у времену попут бројања аута на цести, кретања молекула гаса или течности, раста популације бактерија, обради сигнала. У једној популацији има n јединки са вероватноћом p_n, а број житеља се у произвољном тренутку мења највише за $+1$ са вероватноћом a_n или за -1 са вероватноћом b_n. Постављамо питање како довести у везу низове a_n, b_n, p_n. Овде се не подразумева независност, коју смо раније разматрали [64. Без памћења], па ови бројеви могу бити променљиви.

Претпоставимо да канал преноси податке састављене од слова (лат. littera) азбуке из низа π_0, π_1, ..., π_l да слово π_i због специфичних сметњи при преносу може постати само једно од њему „блиских“ слова $\pi_{i-1}, \pi_i, \pi_{i+1}$. Дакле, претпоставка је да су сви коефицијенти матрице канала $k_{ij} = 0$, ако је $|i-j| > 1$. За такав канал може се рећи да има сепарабилне сметње. Овде су:

$$k_{i,i+1} = a_i, \quad k_{i.i-1} = b_i, \quad (b_0 = 0, \;\; i = 0, 1, \ldots, l)$$

једине вероватноће које могу бити различите од нуле. Пропустимо ли сигнал $\pi_i \in \{\pi_0, \pi_1, \ldots, \pi_l\}$ кроз дугу серију таквих канала, онда је p_j вероватноћа да ћемо на излазу из једног фиксног места серије имати сигнал π_k.

Теорема 1. Нека систем има низ стања $\Pi_0, \Pi_1, \ldots, \Pi_l$, а вероватноћа непосредног преласка система из стања Π_i у Π_j је нула ако је $|i-j| > 1$. Уведимо следеће ознаке:

$$p_n = \Pr(\Pi_n), \quad a_n = \Pr(\Pi_{n+1}|\Pi_n), \quad b_n = \Pr(\Pi_{n-1}|\Pi_n), \quad n = 0, 1, 2, \ldots$$

$(b_0 = 0)$. Означили смо са s најмањи природан број за који је $a_l = b_{l+1} = 0$. Тада је:

$$p_n = \frac{\prod_{i=1}^{n} \frac{a_{i-1}}{b_i}}{1 + \sum_{j=1}^{n} \prod_{i=1}^{j} \frac{a_{i-1}}{b_i}}.$$

Доказ: Према формули потпуне вероватноће имамо:

$$p_n = \Pr(\Pi_{n-1}|\Pi_n)\, p_n + \Pr(\Pi_{n+1}|\Pi_n)\, p_n + \Pr(\Pi_n|\Pi_n)\, p_n$$

$$p_n = \Pr(\Pi_n|\Pi_{n-1})\, p_{n-1} + \Pr(\Pi_n|\Pi_{n+1})\, p_{n+1} + \Pr(\Pi_n|\Pi_n)\, p_n$$

где је $\Pr(\Pi_n|\Pi_n)$ вероватноћа да ће систем остати у истом n-том стању $(n = 1, 2, \ldots)$. Изједначавање десних страна даје релацију

$$(a_n + b_n)p_n = a_{n-1}p_{n-1} + b_{n+1}p_{n+1}$$

где је на левој страни вероватноћа да ће систем напустити n-то стање, а на десној вероватноћа да ће из суседних стања систем ући у n-то. Према томе, вероватноћа да ће систем ући у n-то стање и

прећи у $(n+1)$-о једнака вероватноћи да ће систем ући у $(n+1)$-о стање и прећи у n-то. Заиста, дефинишемо ли случајну променљиву

$$c_n = -a_n p_n + b_{n+1} p_{n+1}$$

лако добијамо ($c_l = 0$) $c_{n+1} = c_n$, па је $c_n = 0$ за свако $n = 1,2,\dots$. Тада је:

$$p_n = \frac{a_{n-1}}{b_n} p_{n-1} = \prod_{i=1}^{n} \frac{a_{i-1}}{b_i} \cdot p_0.$$

Због услова $\sum_{n=0}^{\infty} p_n = 1$ добијамо

$$p_0 = \frac{1}{1 + \sum_{j=1}^{s} \prod_{i=1}^{j} \frac{a_{i-1}}{b_i}},$$

а због $a_l = 0$ је теорема доказана. ■

Сада видимо зашто је ово теорема о сепарабилности, али не и о независности случајних догађаја. Ако бисмо имали и независност, дакле нове исходе увек исте вероватноће независно од прошлих, онда би сви a_i и b_j били константе и посебно било би $\frac{a_{i-1}}{b_i} = q$ константно. Тада се ова расподела своди не неки од претходно разматраних експоненцијалних процеса [64.1. Теорема 1].

74.1. Гравитација

Када је поменути количник q константан, обзиром да је $q = a_{n-1}\!: b_n = p_{n+1}\!: p_n$, интересантнија је ситуација опадања вероватноће са порастом индекса. Тада $p_{n+1} < p_n$ за свако $n = 1,2,\dots$, односно $0 < q < 1$, па имамо:

$$p_0 = \frac{1}{1 + q + q^2 + \cdots + q^l} = \frac{1-q}{1-q^{l+1}} \to 1 - q, \quad l \to \infty,$$

$$p_n = p_0 q^n = p_0 e^{-\lambda n} \quad (\lambda > 0).$$

То је експонекцијална расподела која може интерпретирати једну од сфера вероватноће од раније помињаних [38. Виртуелне сфере]. Раст вероватноћа ка центрима сфера, наравно, може задржати сепарабилност без независности, када коефицијент q не би био константан, а са сферама које тада „имају памћење“.

Пример 1. Посматрајмо вероватноћу $p = e^{-\lambda n}$ као експоненцијалну функцију неке информације $\lambda n = -\ln p$. Променом параметра ($\lambda = \ln b$) мења се база експоненцијалне функције ($p = b^{-n}$), а можемо рећи и јединица информције. Дамо ли том параметру димензију физичког времена, тада други фактор експонента има димензију енергије или потенцијала, јер су информација и дејство еквиваленти.

Свеједно, већа вероватноћа значи мању неизвесност, мању емисију информације, мање случајних догађаја и спорији ток времена. Ово последње једна је од основних теза „теорије информације“, у којој време меримо променама, а основа свих промена су исходи (случајних) догађаја. ❑

Пример говори о потврди методе коју сам сада већ давно описивао (Простор-Време, Економски институт Бања Лука, 2017), о добијању Ајнштајнових резултата гравитације посматрајући слободан пад без очигледне употребе принципа опште релативности.

У том прилогу израчунава се маса m тела које слободно пада у гравитационо поље планете масе M на удаљености r од центра привлачења. Потенцијална енергија прелази у кинетичку енергију тела и додаје се укупној енергији (E). Са порастом брзине падања телу расте маса ($E = mc^2$). Те промене су експоненцијалне ($m = m_0 e^{GM/rc^2}$, m_0 је маса ван гравитације). Оне затим утичу како на трајање релативне јединице времена ($\Delta t = \Delta t_0 e^{-GM/rc^2}$), тако и на радијалне дужине (у правцу кретања). Резултат је да и време успорава, а да се радијалне дужине скраћују на начин предвиђан раније од Ајнштајна (1916).

Са друге стране, тела која падају немају потенцијалну већ кинетичку енергију (кретања) и одозго гледан фотон се понаша као да потиче са извора светлости који се све брже удаљава. Због тога му је опажена фреквенција све мања, а онда то можемо разумети као продужавање трајање периода времена. Време унутар јачег поља тече спорије. Две групе резултата једнаке су, мада други начин избегава јасну употребу принципа опште релативности са којима су првобитно добијени.

Овде видимо и трећи метод са једнаким резултатима. Због веће извесности (у срединама већих вероватноћа) информације су мање, мањи је „проток“ (број или количина исхода) случајности, те је и временски ток спорији. Спорији ток времена гравитационо је привлачан, у своје време уочио је и Ајнштајн, а онда следе и остали релативистички ефекти гравитације. Сада додајмо да спорији ток времена значи више (релативне) вероватноће и мање информације, а њихова привлачност је просто последица „сила неизвесности“, односно принципа минимализма комуникације. Дакле, то је појава општија и од опште релативности.

Сепарабилност фермиона[8] гаранција је важења 1. теореме што се тиче расподеле вероватноћа, а новост је да би гравитација могла бити грана теорије вероватноће. Међутим, та теорема не тражи константност количиника q и, према томе, не подразумева независност простора гравитације, тј. дозвољава и могућност да такав „простор памти“. Обзиром да је „вероватноћа да ће систем ући у n-то стање и прећи у $(n+1)$-о једнака вероватноћи да ће систем ући у $(n+1)$-о стање и прећи у n-то“ (цитат из доказа теореме), за веће извесности наведених вероватноћа рашће константност јачине поља.

74.2. Телефонска централа

Телефонска централа има $l = 1, 2, \dots$ телефонских линија. Позив који стигне када је свих l линија заузето пропада. Означимо ли са Π_n стање централе са тачно n линија заузетих тада теорема има следећу интерпретацију:

- p_n – вероватноћа да је $n = 0, 1, 2, \dots, l$ линија заузето;
- a_n – вероватноћа новог позива када је n линија заузето;
- b_n – вероватноћа да ће један од n разговора завршити.

Претпоставимо да су ове вероватноће (p_n, a_n и b_n) дате у односу на одређени временски период $\Delta t = t_2 - t_1 > 0$ који је врло кратак. Током доказа теореме стоји да је „вероватноћа да ће систем

[8] Fermions, half-integer spin particles, https://www.academia.edu/9994343

ући у n-то стање и отићи у $(n+1)$-о једнака вероватноћи да ће систем прећи у $(n+1)$-о стање и вратити се у n-то", што значи да је за веће извесности наведених вероватноћа број заузетих линија више уједначен, централа тада тежи осцилирању око фиксираног броја позива.

Када вероватноћа новог позива не зависи од броја (n) заузетих линија централе нити од избора почетног тренутка, тада она задовољава услове за експоненцијалну расподелу, тј.

$$a_n(\Delta t) = 1 - e^{-\alpha \Delta t}, \quad \alpha > 0.$$

У граничном случају, када $\Delta t \to 0$, тада $a_n(\Delta t)/\Delta t \to \alpha$.

Са друге стране, када вероватноћа завршетка произвољног једног од (n) текућих разговора не зависи од броја n, нити од избора почетног тренутка (t_1), већ само од трајања разговора Δt, тада је она такође експоненцијална

$$1 - e^{-\beta \Delta t}, \quad \beta > 0.$$

Вероватноћа да неће доћи до прекида нити једног од n оваквих разговора за време Δt је $e^{-n\beta \Delta t}$, па је вероватноћа да ће бити прекинут најмање један

$$b_n(\Delta t) = 1 - e^{-n\beta \Delta t}.$$

Када $\Delta t \to 0$ тада $b_n(\Delta t)/\Delta t \to n\beta$, па коефицијент $a_{n-1}/b_n \to \alpha/n\beta$. Користећи последњу теорему добијамо познате Ерлангове формуле за телефонску централу

$$p_n = \frac{\frac{1}{n!}\left(\frac{\alpha}{\beta}\right)^n}{\sum_{k=0}^{l} \frac{1}{k!}\left(\frac{\alpha}{\beta}\right)^n}, \quad n = 0, 1, 2, \dots, l.$$

Ова вероватноћа p_n не зависи од времена Δt.

Пример 2. Када у току једног часа централа прими 60 позива, просечно је чекање позива 1 min. Ако просечни разговор траје 2 минуте онда се у року једног часа заврши 30 разговора. Тада имамо количник $\frac{\alpha}{\beta} = 2$, па за централу са две линије $(l = 2)$ је $p_0 = 0.2$, $p_1 = 0.4$ и $p_2 = 0.4$. ❑

***

Мултипроцесирање[9] у електронској обради података настаје када више корисника у исто време треба услугу једног рачунара. Централна јединица (CPU) пословних рачунара има m процесора где је обично $m = 1$ или је $m = 2$, док је број захтева далеко већи. Када у току од на пример неколико чаосва захтеви за обрадом података (програми) почињу независно од броја већ пристиглих, као и од избора почетног тренутка, тада имамо ситуацију формално сличну телефонској централи али са једном битном разликом. За број активних програма $n > m$ нови захтеви не пропадају већ иду у тзв. ред чекања до броја l. Међутим, вероватноће a_n и b_n имају сличну интерпретацију.

[9] Р. Вуковић: „Математичка теорија информације и комуникације", Друштво математичара Републике Српске, Бања Лука 1995.

75. Физичка сила

Идеја да се физичке силе представљају (делом) теоријом вероватноће хипотеза је која се из (моје) теорије информације чини веома привлачном. Пре свега, јер је полазиште ове теорије објективна случајност, мање-више присутна у појавама око нас, а вероватноћа и информација мере је тако да више прве значи мање друге. Стога прећутно прихватано начело максимализма, да су вероватнији исходи чешћи, постаје ново начело минимализма: да природа штеди емитовање информације.

Већ са ова два „минимакс начела“, постаје актуелно питање одбојне „силе неизвесности“ [27.], а са њоме и привлачне „силе извесности“ [38. Виртуелне сфере]. Може се показати да за апстрактне просторе вероватноћа, метрички или векторски, важи поопштен Кеплеров други закон: потег од централне константне „силе“ до „набоја“ којег она „покреће“, у једнаким „временима“ пребрише једнаке „површине“. Такође да се наводни набоји тада крећу по коникама (хиперболама), а да је брзина ширења дејства брзина светлости.

Замишљамо даље, посредно, да еквивалентност ових апстрактних и физички реалних сила можда заменимо јединственим третманом обе. Касније, да их заједно са евентуалним будућим налазима третирамо са неке друге стране, сада нама непознатим њима заједничким узроком. Нове теорије отварају нове могућности.

На пример, мултипроцесирање са својим формализмом заузимања n од могућих m канала токова информација, при чему се вишак не већи од l држи у резерви ($m < l$), могуће је пренети на појаве „гравитације“. Вишак пристиглих опција (информације) иде у „ред чекања“, који бисмо тада могли називати додатним димензијама времена, иначе „нормалним“ (Dimensions) овој теорији. Прошли наслов [74.] обрађује сепарабилне процесе аналогне фермионима, рецимо честицама твари, које се не могу наћи у истом квантном стању (Pauli exclusion principle). Исти се дотиче експоненцијалне расподеле вероватноћа [64.1. Независност] која има везе са гравитацијом, поновићу како.

Посматрамо вертикалан пад (Простор-време, стр. 58) тела масе m, енергије $E = mc^2$, ка центру гравитационог поља масе M, на инфинитезималном делу пута r. Рад силе на путу даје:

$$dE = Fdr,$$

$$dmc^2 = -G\frac{Mm}{r^2}dr,$$

$$m = m_0 e^{GM/rc^2},$$

где би маса тела које пада била m_0 у одсуству гравитације, c је брзина светлости у вакууму, а G је универзална гравитациона константа. Кинетичка енергија кретања током слободног падања тачно (потпуно) је заменила потенцијалну гравитације, па их у том смислу сматрамо еквивалентним.

Два тела која једнако падају не осећају гравитациони потенцијал, могуће је сматрати да они и нису у гравитационом пољу. Изводимо закључак да би, са велике висине виђен, фотон унутар тог поља био замењив фотоном извора који се истом брзином v удаљава, једнаком брзини датог тела датог тренутка чија кинетичка енергија, $E_k = mv^2/2$, је једнака поменутој потенцијалној. Због „црвеног помака“ (Доплеров ефекат), фреквенција фотона је мања, трајање осцилација је дуже, а оне мере јединице времена гравитационог поља.

На описани начин разумемо да се маса и енергија тела у гравитацином пољу релативно повећава, да му време успорава, а затим и да радијалне дужине бивају краће, на начин изложен у прилогу. У истом се развоју даље појављује Шварцшилдова метрика

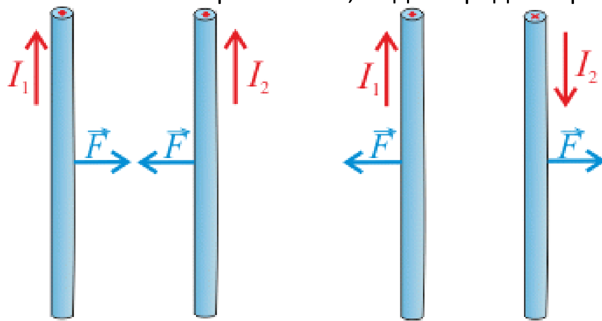
$$ds^2 = \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 + r^2 \sin^2\theta \, d\varphi^2 + r^2 d\theta^2 - \left(1 - \frac{2GM}{rc^2}\right) c^2 dt^2$$

засновано на апроксимацији развоја експоненцијалне функције у ред.

75.1. Елементи

Ове резултате слободног пада у Њутновом гравитационом пољу узимамо озбиљно у преносу на Ајнштајново релативистичко, при чему примећујемо да постоје разлике у јаком и све јачем пољу. Та прва апроксимација, слабог (Њутновог) поља, постаје експоненцијална расподела вероватноћа, што долази са независношћу будућих исхода од претходних, односно да га чини простор који „не памти“, који својом прошлошћу не утиче на сопствену будућност. Релативистичко прецизирање, које се односи на јача поља, доноси ову „меморију“ и њен утицај, односно повлачење простора за супстанцом која се њиме креће.

Ако сила мења вероватноће, онда то ради и промена брзине. То такође приказује Кулонова сила,



односно ток електрона кроз блиске паралелне водиче, на слици лево. Писао сам о томе као о занимљивој хипотези раније (Current), коју сад само понављам уз мале додатке.

Два паралелна електрична проводника, струја у супротним смеровима, индуковеће магнетно поље које их одбија, а поље је привлачно ако су струје у истом смеру. То је познато одавно и пуно од овога што ћу рећи је у електромагнетизму одавно познато.

Сила F_0 по јединици дужине између две паралелне струје I_1 и I_2 раздвојене растојањем r износи

$$F_0 = \frac{\mu_0 I_1 I_2}{2\pi r},$$

где је $\mu_0 = 4\pi \times 10^{-7}$ Tm/A тачно. На тај начин се дефинише да је један ампер струје која кроз сваки од два паралелна проводника бесконачне дужине, раздвојених за један метар у празном простору без других магнетних поља, изазива силу од тачно 2×10^{-7} N/m на сваки проводник.

Ово комбиновано са начелним минимализмом емисије информације доводи до закључка да исти набоји у истосмерном кретању са порастом брзине смањују своје укупне опције, или снагу емисије информације здруженог система, појачавајући узајамну привлачност. Обрнуто чине разноимена наелектрисања у истим условима. Она повећавају своје укупне опције стварајући средину са више емисија информације и тиме појачавају узајамну одбојност.

Гледајући у оквиру кинематике, два сама електрона у узајамном мировању привлачила би се. Они би постали два електрона у супротном кретању и одбојном узајамном силом која би даље падала, али без мењања смера. Међутим, у проводнику са много других електрона, електрон би стјешњен

осциловао у месту, по правцу кретања струје. Сличан првом је случај два „слепљена“ електрона у потпуном узајамном мировању, али такво стање некретања квантног система није могуће.

Теорија информације предвиђа бесконачне (Infinity II) могућности сличних оваквих разлагања, али опет дискретну појаву слободних, физичких информација (Packages). Као „позитивно и негативно“ наелектрисање које се сабира у нулто, или „три основне боје“ које се сабирају у неутралну. То јер постоје низови матрица чији производ даје јединичну $\hat{M}_1\hat{M}_2 \cdots \hat{M}_n = \hat{I}$, или композиције процеса са резултатом почетном вредношћу. Међутим, информације које су покретне, које физичка тела могу размењивати ограничене су еквиваленцијом са физичким дејством. Оно је дно информације стварности, као што су рецимо атоми и молекуле дно хемије, односно молекуларне физике.

Када два проводника нису паралелни него граде неки угао, тада силу F_0 треба још помножити са синусом тог угла, дакле умањити због удаљавања упоредних електрона. То је познато из класичне електродинамике, а овде разумемо као повећање количине неизвесности и њихове одбојне силе. Тиме долази до мешања, до изградње смеша компоненти које су све више делови сложеног, пре него елементарног света физике.

75.2. Инерција

Гравитациона сила не делује на елементарне честице, на фермионе спина $\pm\frac{1}{2}$ и бозоне спина не даљег од ± 1, јер гравитони, претпостављене честице који преносе њене интеракције, имају спин ± 2, а приликом свих интеракција одржава се укупна количина спина. За спин, као и за енергију, импулс, или информацију важе одговарајући закони одржања.

Елементарне честице удружене у атоме и молекуле предмет су комуникације са гравитонима, а такве су смеша оних различитости које су важне одвојене за електромагнетно деловање. Али тада до изражаја долазе друге временске димензије (Dimensions) и заобилажења (Bypass) која се путем њих могу дешавати. Гравитационе честице имају способност „изостајања“ из ове, наше реалности и тиме присуства своје информације у њој, продирући у 6-дим универзум, тачније речено градећи га, више што их је више. Мањак информације такве њихове средине привлачан је, универзално за сву супстанцу, што претпостављам да је „гравитациона сила“.

Када А може непосредно комуницирати са Б казаћемо да су А и Б непосредно узајамно реални. А ако то није могуће, али Б може комуницирати са Ц, онда су А и Ц само (узајамно) реални. У таквој структуру, реалност је више ствар масовности комуникације него изолованих случајева[10]. Поједине случајеве одласка у „друге опције“ зато посматрамо као одласке у псеудо реалност, у паралелну реалност, или паралелне светове. Лако је проверити да је овакав (нови) третман „реалности“ иде уз додатне „временске димензије“.

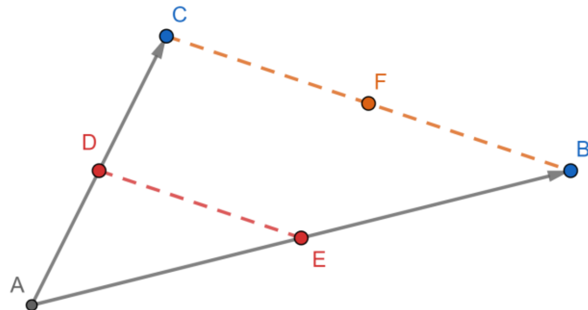
Мањак информације у окружењу масе честица супстанце углавном значи и мањак комуникација међу њима и, према томе, мањак реализација случајних догађаја у односу на посматраче изван. То значи и релативно спорије протицање времена, јер обим догађаја дефинише временски ток, бар што се тиче (моје) теорије информације. Са друге стране, ово „потискивање“ информације у мноштву потсећа на мултипроцесирање са све дужим редом чекања (могу се формално слично

[10] Universe Is Not Locally Real: https://www.scientificamerican.com/article/the-universe-is-not-locally-real-and-the-physics-nobel-prize-winners-proved-it/

третирати), као и на „самоконтролу“ супстанце организоване у атоме, молекуле и смеше. Такође на форме смисленог текста (који има мању информацију од одговарајућег шума), реда уопште у односу на неред.

Већа маса заузима већу запремину, а тада је светлости потребно све више времена да стигне са краја на крај, па су делови истог тела само парцијално истовремени. Такви су квантно спрегнути што их са поменутом компактношћу чини неком „целином“. Утегнутост све веће ове целине све више проблематизује промену њене брзине, због чега су већа тела масивнија, инертнија. Уз то, простирање већих тела више у временске димензије, укључујући прошлост, учиниће да много масивна имају неку сенку коју гравитационо вуку за собом, сенке из прошлости гравитационог деловања на садашњост. Ова (хипо)теза је оно „памћење простора“ које ми је ишло уз објашњење дела загонетне „тамне материје“ у астрономију опажене (израчунате).

Детаљ из „масовне спрегнутости“ приказан је на слици десно. Тачке $A, B, \dots, F$ су догађаји физичке реалности, неког простор-времена. Из тачке A, истовремено излећу тачке B и C у различитим смеровима и брзинама. Нека су D и E средње тачке дужи AC и AB редом, а F средина дужи BC. За D тачке A и C су истовремене, исто тако A и B истовремене су са становишта тачке E, а за неког вањског посматрача нити један пар од низа ових тачака не мора бити истовремен.



Када се деси промена на A, истовремено и редом за E и D дешавају се промене на B и C, иако се неком вањском посматрачу тачке B и C неће појавити као истовремене (такође нити E и D). Тако настаје познато „фантомско деловање“ између пара $A - B$, као и $A - C$, или квантна спрегнутост али и $B - C$ које се тада јавља као „повлачење“ за целином. Ово последње настаје због „утезања“ са осталим тачкама масе. Теорија информације дозвољава постојање и посматрача F за којег су B и C истовремене, који не мора бити у истој реалности са тачкама A, B и C неког даљег посматрача.

Подсећам, слабост деловања гравитације из прошлости на садашњост веома је слаба (системима мера физике које су наш стандард), због временске дужине ict чији се иначе тек квадрат ($i^2 = -1$) појављује у метрикама и због огромне вредности брзине светлости ($c \approx 300\ 000$ km/s). Отуда ће галаксије мање специфичне густине имати мању па и никакву „гравитациону сенку“ (AGC 114905). Зато је повлачење простора за планетом Меркур приметно, довољно значајно да покреће његов перихел у смеру око Сунца, јер је Сунцу та планета најближа и креће је у најјачој гравитацији.

76. Еволуција извесности

Принцип минимализма информације води природна стања у мање информативна, у вероватнија. Једва приметно, али упорно, системи случајности развијају се у извесније. Васиона као каква врста живота на земљи еволуира. Она се стално мења никада не бивајући иста, претварајући се у разне облике, стања и начине, као да не зна шта би са својом неуништивом неизвесношћу, да би током дужих времена ипак остајала са све мање супстанце и више простора, све мањих исказаних и све више потиснутих информација (количина неизвесности).

Постулирање непредвидљивости на начин „теорије информације“ подразумева непрестани развој (Growing). Штавише, васиона је у неком тихим и сталним стањима креације којима не руши законе одржања, али успева мењати и саме своје промене. То је нови моменат, откриће за науку уопште. Неизвесност се цеди емитовањем информације, али се њена укупност одржава растућим обимом посебности, разноврсности.

Пример 1. Након сортирања у опадајући поредак, вероватноће неке расподеле су $p_1, p_2, \dots, p_n$, где је $p_i \ge p_j$ ако $i < j$. Подразумевамо да је свако $p_i > 0$ и, наравно, збир свих $\sum_{i=1}^{n} p_i = 1$. Међутим, што је брже опадање датих вероватноћа мања је Шенонова информација

$$S_{\text{н}} = -\sum_{i=1}^{n} p_i \ln p_i.$$

Са друге стране, држећи се било које функције (од праве линије униформне расподеле, преко експоненцијалне форме до још стрмијих облика), повећањем броја исхода (n) расте Шенонова информација ($S_n < S_{n+1}$). ❑

Другим речима, могућ је општи раст извесности система и уједно очување количине неизвесности уз помоћ повећања броја опција. Раније је поменуто да је ово могуће и „претапањем“ супстанце у простор. Изгледа да се обе ове промене и дешавају.

Пример „умножавања могућности“ је Хигсов (Higgs mechanism) настанак масе стандардног модела физике честица и његова примена у електрослабој теорији и спонтаног рушења симетрије у време ране васионе. Уопште је то и настајање галаксија, планета, па и живота на Земљи. Пример прелаза супстанце у простор (можда) је ширење свемира.

Због саме необичне поставке ове теорије, о објективности случајности, много је наизглед лабавих закључака даље изречено да бих додавао нове. Ипак ћу поменути ширење космоса све даље, како у просторима тако и временима, његово стално настајање, те успоравања времена „садашњости“ (због све мање догађаја, емисија информација). Вредна је помена и ергодичка теорема [61.2.], где је у овом случају њена претпоставка вероватно испуњена (нема сто одстотног тачног преноса било које поруке), а зато и последица, да после много времена (милијарди година трајања васионе) не можемо тачно знати шта се заправо дешавало.

Што су извесности васионе веће, њена сећања су трајнија и паралелне димензије времена дубље. Већи је и сам простор тих псеудо реалности које настају инициране, или коинцидиране, са нашом, тако да попут неодређености микро света, са растом неодређености псеудо космоса упоредно са нашим (произвољним) повећавају се одређености њихових импулса (енергија). Изложена теорија предвиђа промене, нестајање, али углавном настајање, самих закона света.

76.1. Делта функција

Занимљив екстремни случај „расподеле вероватноће“ је функција густине коју је у квантну механику увео П. А. М. Дирак, професор Универзитета у Кембриџу, Енглеска, почетком 20. века разрешавајући неке друге проблеме тадашње физике.

Дефиниција 1. Густина вероватноће $\delta(x)$ дефинисана са:

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases}, \quad \int_{-\infty}^{+\infty} \delta(x)\, dx = 1,$$

назива се Диракова делта функција. ▨

Да појаснимо везу између ове функције, густине вероватноће и преноса информације присетимо се једначине преноса података звуком, електричном струјом, или електро магнетним таласима са каналом континуума $K: \varphi \to \psi$, помоћу линеарног оператора

$$\psi(y) = \int_{-\infty}^{+\infty} k(x, y)\, \varphi(x) dx.$$

Ту је $\varphi(x)$ улазна расподела порука у канал, а $\psi(y)$ излазна. Када је $\psi(y) = \varphi(y)$ за свако y тада је φ својствена функција оператора K.

При томе постоје врло тачни преносници непрекидних података који се на малим растојањима понашају скоро као идентични оператор, тј, такав линеарни оператор K који за сваку густину расподеле $\varphi(x)$ даје $K(\varphi) = \varphi$. Пример таквог оператора за дискретне расподеле је јединична матрица $\hat{I} = [\![(\delta_{ij})]\!]$, док би код континуума то требала бити функција $\delta(x, y)$ за коју важи

$$\varphi(y) = \int_{-\infty}^{+\infty} \delta(x, y)\, \varphi(x) dx$$

баш за сваку расподелу $\varphi(x, y)$. Ма како ови услови изгледали природни, у математичкој теорији показивали су се тешко спојиви.

Пример 2. Нека је $\varphi(x)$ густина униформне расподеле

$$\varphi(x) = \begin{cases} \frac{1}{\varepsilon}, & x \in (0, \varepsilon) \\ 0, & x \notin (0, \varepsilon) \end{cases}$$

где је $\varepsilon > 0$. Тада је

$$\int_{-\infty}^{+\infty} \delta(x, y)\, \varphi(x) dx = \frac{1}{\varepsilon} \int_{0}^{\varepsilon} \delta(x, y)\, dx,$$

а овај интеграл би морао бити једнак $\varphi(x)$ за свако произвољно мало $\varepsilon > 0$. Међутим, тако дефинисана функција $\delta(x, y)$ није интеграбилна нити у Римановом нити у Лебеговом смислу. ❑

Проблем се разрешава помоћу низа функција $f_n(x)$, за $n = 1, 2, 3, \ldots$, који за $n \to \infty$ тежи $\delta(x)$, тој Дираковој делта функцији. Даље се претпоставља за горњи интеграл у Дираковом смислу да је једнак Лебеговом над свим Лебег-интеграбилним функцијама.

Ако је горње претпоставке могуће испунити, тада је за сваку густину расподеле $\varphi(x)$:

$$D(\varphi) = \int_{-\infty}^{+\infty} \delta(x-y)\,\varphi(x)dx = \int_{-\infty}^{+\infty} \delta(z)\,\varphi(z+y)dz =$$

$$= \varphi(y) \int_{-\infty}^{+\infty} \delta(z)\,dz = \varphi(y),$$

а то је управо горња тражена функција.

Другим речима, свака непрекидна функција φ својствена је Дираковом „јединичном оператору“, једнако као што је сваки вектор $\vec{v}$ својствени вектор јединичне матрице, $\hat{I}\vec{v} = \vec{v}$. Са друге стране, Диракову делта функцију можемо протумачити као гранични случај низа $f_1(x, \sigma_1), f_2(x, \sigma_2), \ldots$ који представља низ случајних промењљивих са очекивањем нула и густинама расподеле вероватноће за чије дисперзије важи $\sigma_n \to 0$, када $n \to \infty$. При томе је:

$$S = -\int_{-\infty}^{+\infty} \delta(x) \ln \delta(x)\,dx = 0.$$

Дакле Шенонова информација Диракове функције је нула, што показује да ова „густина“ није нека распршена вероватноћа, већ пре извесност места $x = 0$.

76.2. Феномен величине

Захваљујући „силама неизвесности“, сматрам, чешће се дешавају вероватнији исходи и природа своја стања тежи превести у мање информативна. Мање информативна стања су извеснија, више организована, ефикаснија, мање витална. Како сваки део супстанце има неку „количину опција“, следи закључак да више делова носи више информације. То је парадокс са начелом минимализма, а из њега је излаз „бежање“ масе информације у друге димензије, видели смо. Међутим, ту пристиже у „помоћ“ и закон великих бројева.

Теорема 1. (Чебишевљева неједнакост) Ако случајна промењљива X има варијансу $\sigma^2 = EX^2$, тада за свако $\varepsilon > 0$ важи неједнакост

$$\Pr(|X| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Доказ: Дефинишемо случајну промењљиву

$$Y = \begin{cases} 0, & |X| < \varepsilon \\ \varepsilon, & |X| \geq \varepsilon \end{cases}$$

Тада је $Y \leq |X|$, тј. $Y^2 \leq X^2$ па је очекивање $EX^2 \geq EY^2 = \varepsilon^2 \Pr\,(|X| \geq \varepsilon)$, одакле непосредно следи Чебишевљева неједнакост. ∎

Што је већи број ε то је мања шанса да случајна променљива X узме његову вредност, односно „силе неизвесности“ нагомилаваће исходе унутар дисперзије σ. У случају све већег понављања датог опита проређеност већих одступања биваће све очигледнија. На тај начин ова неједнакост води у тзв. Чебишевљев закон великих бројева [26. Дисперзије].

Рецимо да понављамо $n = 1,2,3,\dots$ пута исти бинарни опит, вероватноће $p \in (0,1)$ да ће се десити жељени исход, односно вероватноће $q = 1 - p$ да се тај неће десити. Нека се жељени исход десио $m \in \{1, 2, \dots, n\}$ пута, односно да се није десио $n - m$ пута. Тада према закону великих бројева, за $n \to \infty$ статистичка вероватноћа $\frac{m}{n} \to p$, односно $\frac{n-m}{n} \to q$, тј. статистичка вероватноћа тежи броју који предвиђа теорија вероватноће.

То је у овим скриптама (Информатичке Теорије) раније расправљано. У наставку погледајмо како се помоћу Чебишевљеве неједнакости може оправдати Диракова делта функција и пренос података линеарним операторима.

Теорема 2. Нека је дат низ случајних променљивих X_n са варијансама σ_n^2 и математичким очекивањима $\mu_n = 0$. Ако $\sigma_n^2 \to 0$ када $n \to \infty$, тада за густине расподеле f_n тих случајних променљивих важе формуле:

$$\lim_{n\to\infty} \int_{-\infty}^{+\infty} f_n(x)\,dx = 1,$$

$$\lim_{n\to\infty} f_n(x) = \begin{cases} +\infty, & x = 0, \\ 0, & x \neq 0. \end{cases}$$

Доказ: Прва једнакост је непосредна последица дефиниције густине расподеле. Затим, полазимо од Чебишевљеве неједнакости:

$$\int_{|x|\geq\varepsilon} f_n(x)\,dx \leq \frac{\sigma_n^2}{\varepsilon^2}, \quad (\forall\varepsilon > 0) \quad <==> \quad \int_{|x|<\varepsilon} f_n(x)\,dx \geq 1 - \frac{\sigma_n^2}{\varepsilon^2}.$$

Због $\sigma_n^2 \to 0$, постоји такво $\sigma_n^2 < \varepsilon^2$, тј.

$$\int_{|x|<\varepsilon} f_n(x)\,dx \geq 1 - \varepsilon.$$

Нека је $M = M(n,\varepsilon) = \sup_{|x|<\varepsilon} f_n(x)$, тада је:

$$M\varepsilon = \int_{|x|<\varepsilon} M\,dx \geq \int_{|x|<\varepsilon} f_n(x)\,dx \geq 1 - \varepsilon,$$

$$M \geq \frac{1}{\varepsilon} - 1 \to \infty.$$

Тиме је доказано друго тврђење. ■

Бирамо ли по једну $f_\sigma(x,y)$ од поменутог низа функција са довољно малим $\sigma^2 > 0$, где је y произвољан параметар, видимо да за свако $\varepsilon > 0$:

$$\left| \int\limits_{|x|\geq\varepsilon} f_\sigma(x)\,dx - \int\limits_{|x|\geq\varepsilon} \delta(x)\,dx \right| \leq \frac{\sigma^2}{\varepsilon^2} \to 0, \quad \sigma^2 \to 0.$$

Отуда за сваку функцију $\varphi(x)$, непрекидну у околини тачке $x = 0$, која је самим тиме и ограничена у таквој околини

$$\left| \int\limits_{|x|\geq\varepsilon} f_\sigma(x,y)\varphi(x)\,dx - \varphi(y) \right| \to 0.$$

У том смислу пишемо

$$\varphi(y) = \lim_{\sigma\to 0} \int\limits_{-\infty}^{+\infty} f_\sigma(x,y)\varphi(x)\,dx.$$

Дакле, пренос података континуума се може описивати горе наведеним линеарним операторима $\psi(y) = \int_{-\infty}^{+\infty} k(x,y)\,\varphi(x)dx$. Друго, за приближно тачан пренос, или пренос података на мањим удаљеностима, постоје следећи раније наведени оператори типа $\varphi(y) = \int_{-\infty}^{+\infty} \delta(x,y)\,\varphi(x)dx$. Трећа претпоставка нам је потребна да бисмо такав пренос описали као процес, односно као функцију непрекидног параметра t који се појављује у свакој серији канала, или у сваком каналу серије.

Три наведена захтева се своде на следеће експлицитне услове:

a) $k(t,x,y) \geq 0$, за свако t,x,y;
b) $\int_{-\infty}^{+\infty} k(t,x,y) = 1$;
c) $\lim\limits_{t\to 0} \int_{-\infty}^{+\infty} k(t,x,y)\varphi(x)\,dx = \varphi(y)$.

Ако постоји таква функција канала $k(t,x,y)$ можемо је интерпретирати као онај део преносника кроз који за време t улазна порука x изађе као порука y. Општије речено, φ и ψ су произвољне непрекидне расподеле трансформисане линеарним оператором $K_t\colon \varphi \to \psi$, уз услов $K_t\colon \varphi \to \varphi$. Хомоген процес би морао испуњавати и додатни услов

d) $K_{t_1} \times K_{t_2} = K_{t_1+t_2}$ за свако t_1, t_2.

Због услова c) је дисперзија (варијанса) за инфинитезимални део преносника нула, па је практичније имати дисперзију јединичног интервала ($t = 1$) ако је процес хомоген, односно густину дисперзије.

Дакле, несумњиво се пренос података може обављати линеарним операторима, штавише, то је могуће и представљеним непрекидним функцијама. Обзиром да је информација ткиво простора, времена и материје (према претпоставци ове теорије), то и саме бесконачности (мислим и на оне

из теорије скупова) треба некако укључити у њене реалности. Ово су предлози (које можете наћи у мојим текстовима и ранијих година).

Уобичајено дефинисање непрекидне функције $f: D \to \overline{D}$, промењљивих из домена (D) у вредности кодомена ($\overline{D}$), каже да за свако $x_0 \in D$ и свако $\varepsilon > 0$ постоји $\delta > 0$ такво да за све $x \in D$ такво да из $|x - x_0| < \delta$ следи $|f(x) - f(x_0)| < \varepsilon$. Оно сада преводимо у перцепције, да ма како била мала неодређеност положаја постоји субјекат (ситуација) који је може опажати (по цену губитка неких других одређености). Исто је са временом, импулсом, енергијом, уопште обзерваблама.

Један од предлога за кардиналне бројеве (пребројиве и непребројиве бесконачности) је даљи развој идеје непредвидљивости саме информације. Ма како великим космос имали (замишљали), он опет има околину у односу на коју је нека неизвесност. Друго, развој догађаја космоса не само да је преформулисање неких датих особина, већ и њихово ново настајање. Другим речима, оно што се око нас догађа нису само рекреације (прерасподеле) увек исте смеше могућности, већ и макар мале, сталне, упорне креације, уз важење закона одржања количине.

Креације космос добија из околних бесконачности (којима се може стално нешто одузимати а да оне увек остају исте), а своју компактност може одржавати попут виртуелних сфера [38.], унутар којих се веће вероватноће одржавају под притиском вањских већих неизвесности. До даљег не видим у тим (хипо)тезама значајне недоследности и волео бих да ми на њих евентуалне неко укаже што пре.

76.3. Моменти

Нека је кроз канал послан сигнал x који је стигао као сигнал y. Разлика $y - x$ је грешка преноса, а средња вредност такве грешке за јединични интервал времена хомогеног канала је

$$\lim_{t\to 0} \frac{1}{t} \int_{-\infty}^{+\infty} (y - x) k(t, x, y)\, dy = \mu(x).$$

Средње квадратно одступање, или варијанса је

$$\lim_{t\to 0} \frac{1}{t} \int_{-\infty}^{+\infty} (y - x)^2 k(t, x, y)\, dy = \sigma^2(x).$$

Ако процес није хомоген, онда су ови лимеси једноставно густине момената у почетном тренутку. Даље претпоставимо да ови моменти постоје, те да постоје парцијални изводи

$$\frac{\partial k(t,x,y)}{\partial x}, \quad \frac{\partial^2 k(t,x,y)}{\partial x^2}$$

као непрекидне функције за сваки t, x, y. Егзистенцију траженог процеса канала уз наведена три услова можемо доказати позивајући се на познате парцијалне једначине Колмогоровав за непрекидне случајне процесе, од којих је прва

$$\frac{\partial k(t,x,y)}{\partial x} + \mu(x) \frac{\partial k(t,x,y)}{\partial x} + \frac{1}{2}\sigma^2(x) \frac{\partial^2 k(t,x,y)}{\partial x^2} = 0.$$

Доказ ове диференцијалне једначине је у следећем наслову.

Пример 3. Ако су моменти грешке константни, без умањења општости можемо претпоставити да су они $\mu = 0$ и $\sigma = 1$, па се дата једначина Колмогорова своди на једначину провођења топлоте

$$\frac{\partial k}{\partial x} + \frac{1}{2}\frac{\partial^2 k}{\partial x^2} = 0,$$

чије је решење нормални канал

$$k(t, x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}, \qquad t > 0.$$

Да је испуњен услов хомогености d) можемо проверити непосредном интеграцијом

$$k(t_1 + t_2, x, z) = \int_{-\infty}^{+\infty} k(t_1, x, y) k(t_2, x, z)\, dy.$$

Проверите! ❑

Приметимо да је једно од скривених открића овог примера, једначине провођења топлоте у стохастичким процесима Колмогорова, управо то што је, да је пренос осцилација молекула са тела веће на тело мање температуре процес преноса информације (количине могућности). Док се осцилације смирују, информација (обим опција) га напушта, тело се хлади и ентропија му расте. Пад температуре шири се и уситњава исходе (пример 1), ради одржавања укупне информације.

Тако, на пример, чисти гасови се могу одвајити од ваздуха тако што ваздух прво охладимо док не постане течан, затим дестилујемо компоненте на различитим температурама кључања. Тај процес даје гасове велике чистоће. Енергетски је интензиван, а осмислио га је Карл фон Линде већ током 1895. да би био први пут индустријски кориштен након седам година, па све до данас.

Други пример је на следећој слици лево, уз опис „технологије за добијање боје за шаргарепу“[11]. То је фазно одвајање загрејаног (70-75° C) сока од шаргарепе након хлађења на 40° C а затим све до собне температуре, те даље на 3-4° C када се уводи фина мрежа у цилиндар.



Теорија информације даје једно једноставно објашњење ове коинциденције, да се снижавањем температуре олакшавају сепарације, због могућег општег раста извесности система уједно са очувањем количине неизвесности уз помоћ повећања броја опција.

[11] AstanovSalih Husenovich, AbduqodirovAbduahadToshmurodovich, RaupovaIrodaBarakayevna, MakhmudovMaxmudIdrisovich, KasimovaGuzalKarimovna: „*Effective Technology For Obtaining Carrot Dye*“, European Journal of Molecular & Clinical Medicine, ISSN 2515-8260, Volume 08, Issue 01, 2021.

77. Колмогорова процес

Најубедљивије и најопштије што има теорија случајних процеса континуума данас су парцијалне диференцијалне једначине Колмонгорова, од којих ћемо прву извести. Разлог њихове занимљивости нама је једноставна интерпретација као серијског споја дугог низа преносника континуалних порука, а са друге стране јер се оне могу подвести под једначине топлоте. Пренос топлоте процес је попут преноса информације, први кроз пренос већих у мања титрања молекула, а други процес је развој догађаја у вероватније, рецимо „преносом" у стања мање информативна.

Пример 1. Полазећи од дефиниције канала у ужем смислу

$$\eta(z) = \int_{-\infty}^{+\infty} k_2(y,z)\,\psi(y)dy, \quad \psi(y) = \int_{-\infty}^{+\infty} k_1(x,y)\,\varphi(x)dx$$

где сигнал има ток $x \to y \to z$, а густине расподела датих сигнала $\varphi \to \psi \to \eta$, можемо дефинисати серијски спој два канала K_1 и K_2 густином

$$k(x,z) = \int_{-\infty}^{+\infty} k_2(y,z)k_1(x,y)\,dy.$$

Ова је функција ненегативна за свако x, z и нормирана по z, те она јесте добро дефинисана густина канала у ужем смислу. Композицију пар канала је убедљивије проучавати на овај начин. Међутим, када имамо дужу серију канала онда добијамо непрактичне вишеструке интеграле, а тада метода Колмогорова показује своје предности. ❑

Проводник података произвољне дужине схватамо као серијски спој, попут пара функција $k(x,z)$ из примера, али бесконачно много инфинитезималних проводника. Порука ухваћена на произвољном делу проводника је непрекидна случајна променљива $x, y, z, \dots$ са густином расподеле φ, ψ, η, … . Зато мењамо начин приказивања композиције примера.

Претпостављамо да постоји непрекидан параметар t такав да на произвољном али фиксном делу проводника ухваћени сигнал x зависи од t. Тада је $x = x(t)$ или $\xi(t) = \varphi(x,t)$. Дакле, имамо процес у свакој тачки проводника. При томе важе следеће једнакости:

I. $k(x,t_1;y,t_2) \ge 0, \quad \int_{-\infty}^{+\infty} k(x,t_1;y,t_2)\,dy = 1, \quad 0 < t_1 < t_2;$

II. $\lim\limits_{\Delta t\to 0} \frac{1}{\Delta t}\int_{|y-x|\ge\varepsilon} k(x,t;y,t+\Delta t)\,dy = 0, \quad \forall\varepsilon > 0;$

III. $\lim\limits_{\Delta t\to 0} \frac{1}{\Delta t}\int_{|y-x|<\varepsilon} (y-x)\cdot k(x,t;y,t+\Delta t)\,dy = a(x,t);$

IV. $\lim\limits_{\Delta t\to 0} \frac{1}{\Delta t}\int_{|y-x|<\varepsilon} (y-x)^2\cdot k(x,t;y,t+\Delta t)\,dy = b(x,t);$

V. $k(x,t_1;z,t_3) = \int_{-\infty}^{+\infty} k(y,t_2;z,t_3)\cdot k(x,t_1;y,t_2)\,dy.$

Прва једнакост (I) говори о густини (k) расподеле вероватноћа. Нормираност ту значи да је сигнал y у тренутку t_2 имао неку вредност од $-\infty$ до $+\infty$. За довољно дуго време $t_2 - t_1$ сигнал y може постати било шта. Међутим, ако $t_2 - t_1 \to 0$, тада $y \to x$, тј. важи друга једнакост (II). Тај услов (II) изражава чињеницу густине, да вероватноћа промене сигнала x за више од ε тежи нули, када Δt тежи нули. То је захтев континуалности процеса.

Како интервал у једнакости (II) за $|y-x| \geq \varepsilon$ тежи нули, тако његов остатак за $|y-x| < \varepsilon$ тежи 1 због претходне (I). Прецизније, за ове остатке се тражи да униформно конвергирају (по x), а отуда једнакости (III) и (IV).

Први од ових лимеса $a(x,t)$ можемо интерпретирати као густину математичког очекивања промене сигнала x у тренутку t. Када је $x=0$, а канал хомоген онда је тај први лимес (III) очекивање излазног сигнала y након времена $t=1$. Други од ових лимеса (IV) је близак појму варијансе (густине дисперзије).

Прва од функција (I) јединствена је за цели проводник, што је изражено произвољношћу параметра t. Отуда последња од наведених једнакости (V), где на левој страни имамо трансформацију сигнала x из тренутка t_1 у сигнал z у тренутку t_3, док на десној страни имамо то исто преко сигнала y у тренутку t_2. Претпоставимо ли да постоје парцијални изводи:

$$\frac{\partial k(x,t_1;y,t_2)}{\partial x}, \quad \frac{\partial^2 k(x,t_1;y,t_2)}{\partial x^2}$$

и да су оне непрекидне функције за свако $x, t_1, y, t_2 > t_1$, тада вреди следећа једначина.

77.1. Пренос информације

Теорема 1. (Једначина Колмогорова) Функција $k(x,t_1;y,t_2)$ дефинисана горњим релацијама (I-V) је решење једначине

$$\frac{\partial k(x,t_1;y,t_2)}{\partial t_1} + a(x,t_1)\frac{\partial k(x,t_1;y,t_2)}{\partial x} + \frac{b(x,t_1)}{2}\frac{\partial^2 k(x,t_1;y,t_2)}{\partial x^2} = 0.$$

Доказ: Полазећи од последње (V) лако добијамо

$$k(x,t_1-\Delta t;y,t_2) = \int_{-\infty}^{+\infty} k(z,t_1;y,t_2)k(x,t_1-\Delta t;z,t_1)\,dz.$$

Полазећи од прве (I) множењем имамо

$$k(x,t_1;y,t_2) = \int_{-\infty}^{+\infty} k(x,t_1;y,t_2)k(x,t_1-\Delta t;z,t_1)\,dz.$$

Одузимањем последње две једнакости добијамо (*):

$$\frac{k(x,t_1-\Delta t;y,t_2)-k(x,t_1;y,t_2)}{\Delta t} = \frac{1}{\Delta t}\int_{-\infty}^{+\infty}[k(z,t_1;y,t_2)-k(x,t_1;y,t_2)]k(x,t_1-\Delta t;z,t_1)\,dz.$$

Овај интеграл можемо раставити на два сабирка:

S_1 где је област интегрирања $|z-x| \geq \varepsilon$,

S_2 где је област интегрирања $|z-x| < \varepsilon$.

Ако пређемо на лимес када $\Delta t \to 0$, због (II), имамо $S_1 \to 0$, па ће нам остати само други сабирак. Тејлорова формула даје

$$k(z,t_1;y,t_2) = k(x,t_1;y,t_2) + (z-x)\frac{\partial k(x,t_1;y,t_2)}{\partial x} + \frac{(z-x)^2}{2}\frac{\partial^2 k(x,t_1;y,t_2)}{\partial x^2} + o(z-x)^2$$

па је:

$$S_2 = \frac{\partial k(x,t_1;y,t_2)}{\partial x}\frac{1}{\Delta t}\int\limits_{|z-x|<\varepsilon}(z-x)k(x,t_1-\Delta t;z,t_1)\,dz +$$

$$+\frac{\partial^2 k(x,t_1;y,t_2)}{\partial x^2}\frac{1}{2\Delta t}\int\limits_{|z-x|<\varepsilon}[(z-x)^2 - o(z-x)^2]k(x,t_1-\Delta t;z,t_1)\,dz \to$$

$$\to a(x,t_1)\frac{\partial k(x,t_1;y,t_2)}{\partial x} + \frac{b(x,t_1)}{2}\frac{\partial^2 k(x,t_1;y,t_2)}{\partial x^2}.$$

Како је

$$\lim_{\Delta t\to 0}\frac{k(x,t_1-\Delta t;y,t_2) - k(x,t_1;y,t_2)}{\Delta t} = \frac{\partial k(x,t_1;y,t_2)}{\partial t_1}$$

то је тврђење доказано. ■

На сличан начин, полазећи од $t_1 + \Delta t$ уместо $t_1 - \Delta t$, добијамо исти израз на десној страни, попут (*) у доказу. Међутим, на левој страни имамо

$$\lim_{\Delta t\to 0}\frac{k(x,t_1+\Delta t;y,t_2) - k(x,t_1;y,t_2)}{\Delta t} = \frac{\partial k(x,t_1;y,t_2)}{\partial t_1},$$

па доказујемо да важи и једначина

$$\frac{\partial k}{\partial t_1} - a\frac{\partial k}{\partial x} - \frac{b}{2}\frac{\partial^2 k}{\partial x^2} = 0.$$

Она је слична једначини топлоте [76.3. Пример 3].

Пример 2. Када је дисперзија $b(x,t_1)$ велики број, сразмерно у односу на остале сабирке једначине, имамо приближну једнакост

$$\frac{\partial^2 k(x,t_1;y,t_2)}{\partial x^2} = 0.$$

Отуда је $k(x,t_1;y,t_2) = f(y,t) + xg(y,t)$, где су f и g непознате функције. Из услова (I) добијамо $g = 0$, тј. k не зависи од x. Другим речима, канал је црна кутија. ❑

Када дисперзија $b(x,t_1) \to 0$, можемо претпоставити да је очекивање $a(x,t) = 0$, па је $k = 0$ скоро свуда. Због нормираности ова функција мора имати сингуларитет сличан Дираковој делта функцији, па канал постаје друга крајност (примеру 2), тј. бела кутија. Постоји прилично видљива веза ових једначина са једначинама преноса топлоте, аналогија на коју се вреди покушати осврнути. То је тема следећег поднаслова.

77.2. Поређења

Објективна неизвесност је кључна за (моју) „теорију информације“. То је пре свега[12] „објективно“ не познавање неких информација пре њиховог добијања. Не знамо унапред, на пример, који број из лото бубња ће бити извучен, или како се звала особа са којом смо се управо упознали и слично. Та потреба за комуникацијом сведочи о важности информација у свету реалности, али такође и о немогућности поседовања свих знања, затим о Геделовој теореми (ма колико велики скуп тачних тврђења дата теорија имала, увек постоје тачна тврђења која се из њих не не могу извести).

Ма како велики субјекат (комуникације) био, он не може имати сва знања и, према томе, прва од наведених „објективности неизвесности“ универзална је, принципијелна, она сведочи о свеприсутности информације око нас. Оно што је мање извесно мање је информативно, или исто речено другачије, када знамо да ће се нешто десити па се то и деси онда то и није нека вест. Отуда, мање вероватни исходи емитују више информације.

Веће титрање молекула, које садржи више степени слободе, односно које носи већу неизвесност, више потенцијалне информације садржи. Када се та искаже, топлота флуида преноси се (са тела веће температуре на суседно тело мање тепмературе), па сада кажемо да информација титрања прелази на топлијег на хладније. Ентропија тела које емитује информацију повећава се, док му се укупна (потенцијална) информација смањује. То би требало бити одговарајуће објашњење другог закона термодинамике са становишта ове теорије информације.

Према првом закону термодинамике укупна енергија система, учесника у преносу топлоте, остаје константна. То је овде такође случај, када је информација еквивалент дејсту (производу енергије и времена), а јединице времена свих делова система су једнаке. У релативистичким ситуацијама, у релативном систему споријег тока времена (због брзине кретања, или јачег гравитационог поља), онолико пута мањи део интервала времена даће толико пута већу опажену енергију, јер дејства су им једнака.

Дакле, енергија фотона, $E = hf$, који се креће брзином светлости c, зависи само до фреквенције $f = 1/\tau$, где је $\tau = c\lambda$ време једног титраја а λ таласна дужина, према горњем тумачењу, постаје израз $E\tau = h$ константне вредности Планкове h. Честице које се крећу брзином светлости немају сопствено време (време им стоји) и према законима вероватноће налазе се у низу тренутака свог посматрача. Уколико посматрач има масу (m) он има и сопствени ток времена из којег опажа низ положаја светлости као њено кретање.

Уопште честице са масом импулса $p = mv$ зато имају мању брзину од светлости ($v < c$), јер им се титраји (неизвесности) простиру дубље у временске димензије. Оне су зато инертне (објашњење уз слику у поднаслову „75.2. Инерција“) јер су сложено, тј. „масовно спрегнуте“. Било како било, путовање честице простор-временом углавном је највероватнијим путањама, на начине преноса информације каналима. Због опште неизвесности, ма како дати пренос података био тачан он је увек са неким макар веома веома малим грешкама. Зато у овој теорији увек морамо рачунати на ергодичку теорему [61.2].

---

[12] Подразумевам Хајзенбергове релације неизвесности (да тачнијим одређивањем импулса нетачније сазнајемо положај честице, односно да тачнијим одређивањем енергије нетачније сазнајемо тренутак дешавања) и сматрам их тривијалним за овај део излагања.

Због евентуалних грешака преноса информација, из прошлости ка садашњости васионе, старењем догађаји нам постају неизвеснији. Верујемо да је пре 13,8 милијарди година био „велики прасак“ (Big Bang) иза којег „није било свемира“. Доследно ергодичкој теореми, овде даље сматрамо да је тај древни догађај крај наше моћи опажања, али не и „реалност“ каква је „заиста била“. Другим речима, постулирана објективна неизвесност, било да је Хајзенбергових релација неодређености, или на начине горе поменуте, води нас у екстремне космичке неизвесности.

Временска баријера неизвесности (можда 14 милијарди година) васионе, коју сада откривамо, те позната просторна баријера граница видљиве васионе[13], као и субквантне величине мање од нај мањег дејства, објективно су нам границе макар неке извесности. Цеђење, емисија информације из таквих, даје нам понеко сазнање упркос начелу штедње информације. Овај минимализам који нас у информисању „објективно“ спутава еквивалент је „чешћем дешавању вероватнијих исхода“, или „принципу најмањег дејства“ физике, а оба произилазе из одбојне „силе неизвесности“ и ко зна чега још. То су новине ове теорије.

Према томе, оно што видимо као позадинско микро таласно зрачење из самих почетака настанка васионе, због ергодичке теореме, заправо је шум канала преноса информације. Та далека васиона свугде скоро иста својствена је вредност информације која је допутовала до нас, са краја времена до наше садашњости. Шта год да је тамо тада било, ми видимо шум у облику „великог праска“ и верујемо да је након дугог низа копирања нама испоручена информација (или израчуната по нама датим законима физике) тачно једнака оригиналној.

Информације које васиона преноси трајањем мењају се ка својственим дате васионе, а у мањој мери и саме мање честице мењају саму васиону. Додатно, зато што се и сами догађаји васионе помало креирају, поред тога што се рекреирају (дати препакују), ова процењивања прошлости утолико су тежи посао.

Укратко, васиона се хлади, јер све даље (старије) њене делове виђамо као ближе „црној кутији“. А топлота која се „креће ка нама“ сведочи о повећању ентропије супстанце и топљењу њене информације у простор-време којег зато сматрамо има све више. Једнако доследно чини се да би могло бити и следеће тумачење. Супстанце је све мање на уштрб све више простора, а прва се лакше трансформише у другу, па је и све мање исхода, све мање емитованих информација, те је све спорије протицање времена.

Према овом додатном (другом) тумачењу, наша садашњост као да пропада у црну рупу, при чему нам време тече све спорије а дужине постају све краће. Прошлост нам је све даља (у оба смисла, просторном и временском). Због истог, васиона нам се чини све већом, али онда и наше време у односу на све даље галаксија тече све спорије. Део тог успоравања компензује се удаљавањем. Та реалност која нас окружује попут сфера је све већих извесности окружених неизвесношћу, њеним силама истиснута.

Не морају оваква тумачења једна друго оповргавати, напротив, наизглед различите тачне теорије могу једна другу подупирати попут геометрије, алгебре (координатна геометрија) и вероватноће. Овде имамо и уситњавање система којем се парцијалне информације смањују, због конзервације

[13] Даље галаксије све се брже удаљавају, све даља васиона све се брже шири, тако да видимо само онај њен део који се креће до брзине светлости.

количине. То је увећавање различитости при умањивању информације, распадању глобализма[14] у случају евентуалног утезања света бољом организацијом, као кристализација течности приликом хлађења (губитка информације њених молекула) а затим и одвајање. Увећање реда, ефикасности, јасноће текста, значе смањивање информације, а оно се одупире на своје начине, као и покушаји повећања информације рецимо инсистирањем на једнакостима.

Када има где, информација (неизвесност) система кретаће се од веће ка мањој вредности, што је управо спонтано хлађење (повећање ентропије) тела, односно смиривање буре (тежњом честица за мањим физичким дејством), паду флуида избачених из вулкана или гејзира због сила теже. Тако се понаша и развој живота, путем младости преко зрелости и старости до смрти.

Једначина топлоте даје решења функције чији врхови (локални максимуми) постепено опадају, а депресије (локални минимуми) бивају попуњене. Вредност у неком тренутку остају стабилне само док су једнаке просецима свог непосредног окружења. Суптилнија последица је облик максимума чије вредности у било ком региону медијума не прелазе највеће вредности која су се пре десиле, осим ако нису биле на граници. То јест, максимална температура у региону може порасти само ако топлота долази споља. Ово је својство параболичких парцијалних диференцијалних једначина које се лако доказује.

Следеће занимљиво својство је да чак и ако ток топлоте у у почетку има оштар скок вредности преко неке површине унутар медијума, тај се одмах изглађује тренутним, веома кратким и веома брзим протоком топлоте кроз ту површину. На пример, ако се два изолована тела, у почетку на уједначеним али различитим температурама додирују једно друго, температура на контактима ће одмах попримити неку средњу вредност, са зоном варирања постепене промене около.

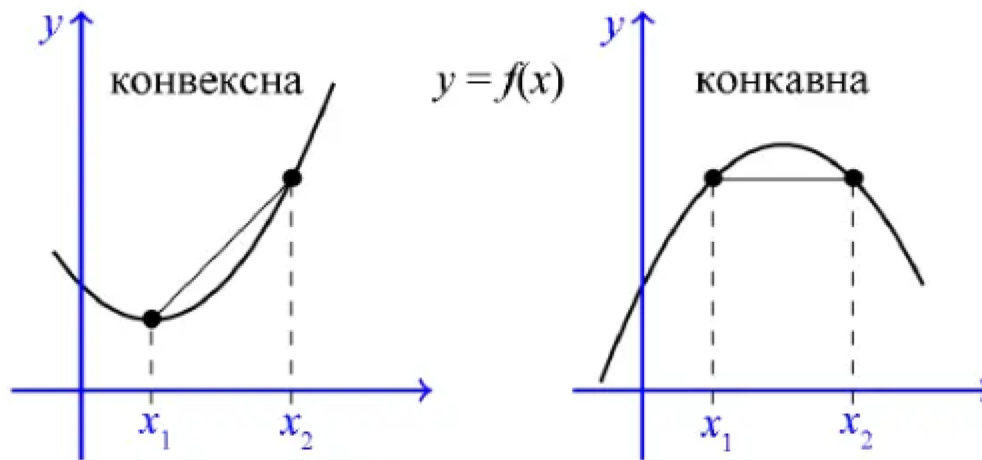
Информација перцепције ($S = ax + by + cz + \cdots$) већа је када се сударају истоврсни укуси, већи са већим а мањи са мањим ($3 \cdot 3 + 2 \cdot 2 + 1 \cdot 1 > 3 \cdot 1 + 2 \cdot 2 + 1 \cdot 3$). Тада су комуникације веће, па је и притисак начелног минимализма информације већи ка разводњавању, организовању, ефикасности, ради каналисања, елиминисања „сувишних“ опција чиме смањивања информације. Али са мањим појединим информацијама увећава се раслојавање целине (закон одржања количине) и процес који личи на кристализацију течности хлађењем и даље пуцање на мање делове, постаје познати нам процес откривања детаља фокусирањем и проучавањем целине.

[14] Глобализам у савременим друштвеним кретањима.

78. Средња вредност

Када имамо неку смешу различитих вредности $x_1, x_2, \dots, x_n$ неједнако заступљених, вероватноћа појављивања редом $p_1, p_2, \dots, p_n$, а разматрамо целину не марећи за делове, тада замишљамо да радимо са неких n једнаких делова вредности $x_0 = (x_1 + x_2 + \dots + x_n)/n$. Често ћемо тада добити задовољавајуће резултате, као да радимо са стварним различитим вредностима.

На пример, уместо са свим деловима тела лакше је радити само са његовим тежиштем, нарочито када рачун даје једнаке резултате. Тако радимо са центрима гравитације небеских тела, уместо са свим честицама које та тела садрже и, као што знамо, у космичко ракетној технологији, довољно тачно израчунавамо трајекторије. Слично опажамо чашу док пијемо воду из ње, не марећи за број и распореде молекула од којих се она састоји. Теорија информације са овом применом има своја занимљива запажања.



На слици, коју сам позајмио из моје збирке задатака за I разред гимназија[15], прва функција је удубљена (конвексна) а друга је испупчена (конкавна). Дуж која спаја две тачке графа лево биће цела изнад графа, а одговарајућа дуж десно испод графа. Када су апсцисе крајњих тачака дужи x_1 и x_2, тада за свако $x \in (x_1, x_2)$ постоје два реална позитивна броја p и q за које важи $p + q = 1$, а да је при томе $x = px_1 + qx_2$.

Тачка криве апсцисе x има висину, тзв. ординату $f(x) = f(px_1 + qx_2)$, а ордината тачке спојнице је $pf(x_1) + qf(x_2)$. Са слике се види да на конвексној и конкавној функцији важе једнакости:

$$f(x) \le pf(x_1) + qf(x_2), \quad f(x) \ge pf(x_1) + qf(x_2),$$

редом, где једнакости важе ако и само ако је и $y = f(x)$ права линија као и спојница. То је иначе добро позната Јенсенова неједнакост коју сам овде доказивао и помоћу логаритама [19.].

Лема 1. За конвексну функцију $y = f(x)$ важи неједнакост

$$f(p_1x_1 + p_2x_2 + p_3x_3) \le p_1f(x_1) + p_2f(x_2) + p_3f(x_3),$$

где су p_1, p_2, p_3 позитивни реални бројеви такви да је $p_1 + p_2 + p_3 = 1$.

[15] Слика 1.19, https://www.academia.edu/20139005/Zbirka_zadataka_I_za_gimnazije

Доказ: Нека је $p_1 \neq 1$. Тада је $1 - p_1 = p_2 + p_3 \neq 0$, па имамо редом:

$$f(p_1x_1 + p_2x_2 + p_3x_3) = f\left[p_1x_1 + (1 - p_1)\frac{p_2x_2 + p_3x_3}{1 - p_1}\right] \leq$$

$$\leq p_1f(x_1) + (1 - p_1)f\left(\frac{p_2x_2 + p_3x_3}{1 - p_1}\right) = p_1f(x_1) + (1 - p_1)f\left(\frac{p_2x_2 + p_3x_3}{p_2 + p_3}\right)$$

$$= p_1f(x_1) + (p_2 + p_3)f\left(\frac{p_2}{p_2 + p_3}x_2 + \frac{p_3}{p_2 + p_3}x_3\right)$$

$$\leq p_1f(x_1) + (p_2 + p_3)\left[\frac{p_2}{p_2 + p_3}f(x_2) + \frac{p_3}{p_2 + p_3}f(x_3)\right]$$

$$= p_1f(x_1) + p_2f(x_2) + p_3f(x_3).$$

А то је и требало доказати. ∎

Из примене претходне неједнакости види се да у овој леми једнакост важи ако и само ако је $f(x)$ функција равне линије. Индукцијом се ова лема лако поопштава на ширу Јенсенову неједнакост

$$f\left(\sum_{k=1}^{n} p_k x_k\right) \leq \sum_{k=1}^{n} p_k f(x_k),$$

која важи за све конвексне функције, за произвољно $n = 1, 2, 3, \ldots$, при чему је $\sum_{k=1}^{n} p_k = 1$ и за свако $k = 1, 2, \ldots, n$ је $p_k > 0$. Приметимо да ова неједнакост постаје једнакост и у случају да је $x_1 = x_2 = \cdots = x_n$. Друго, да обрнута неједнакост „≥“ важи за конкавну функцију, јер ако је $f(x)$ конвексна, онда је функција $y = -f(x)$ конкавна. Ова два случаја говоре о примени са почетка овог наслова, а ево како се то односи на теорију информације.

Логаритамска функција је конкавна. Хартли је њоме одредио информацију X једнако вероватних исхода, са $f(X) = \log X$, где базу логаритма можемо бирати дефинишући тако јединицу мерења информације. Отуда Јенсенова неједнакост за информацију каже да је

$$\log\left(\sum_{k=1}^{n} p_k x_k\right) \geq \sum_{k=1}^{n} p_k \log x_k.$$

На левој страни је информација израчуната према оцењеној средњој вредности (тежишту) целог система, иначе не једнако вероватних исхода, а на десној била би средња информација детаља истог система. Другим речима, удруживањем детаља у систем, средња вредност информације порашће, сем када су ови детаљи једнако информативни (када Јенсенова неједнакост постаје једнакост). Такође, средња информација једнако вероватних исхода максимална је.

Систем ће спонтано тежити раслојавању да смањи информацију (начело минимализма), али ће тада повећати број независних компоненти (закон одржања).

79. Емергенција

У филозофији, теорији система, науци и уметности „емергенција“ (Emergence) долази са појавом својстава неке целине којих њени делови немају сами. Она је појава особина, понашања, дејства делова која се јављају само приликом њиховог удруживања.



На слици је мост[16] који су направили мрави користећи сопствена тела како би ишли пречицом на путу који повезује њихово гнездо са извором хране. Како могу изградити такав мост, с обзиром да имају веома мали мозак? Како се договарају шта мора да се уради, како састављају план за мост, како и када одлучују да га саграде и демонтирају?...

Много је сличних примера појава насталих удруживањем, још од времена Аристотела па до данас, углавном „необјашњивих“ текућој науци и не третираних теоријом информације. Оно што се овде може називати таквом емергенцијом, пре свега је вишак Хартлијеве информације

$$\Delta\mathrm{H} = \log\left(\sum_{k=1}^{n} p_k x_k\right) - \sum_{k=1}^{n} p_k \log x_k \geq 0,$$

где неједнакост постаје једнакост када и у горњој [78.] Јенсеновој неједнакости. Систем који има тај вишак емергенције, видели смо, напетији је и виталнији од опуштеног, јер је вишак количине

[16] Andrea Faré, Mar 27, 2016: What can ants teach us about organization design?

опција, дакле информације и деловања, водећа одлика живих бића. Спонтано системи ће тежити ка развоју у мање информативне (више вероватне), мањим дејствима, мањем отпору.

Код све неживе твари нама данас познате физике имамо управо најмања дејства, највећу могућу опуштеност и препуштање другим силама, која произилазе из решења Ојлер-Лагранжових једначина дефинисаних помоћу тог принципа. Тај минимализма интеракција важи и за олује, гејзире, тела у кретањима под дејствима сила, али не и за жива бића. Међутим и ова потоња, која немају поменуте минимуме, њима теже и зато након успешног развоја младости, стижемо до зрелости, затим старости и смрти.

Друге и детаљније примере информатичког описа емергенција налазе се у мојој недавној књизи[17]. Тема тих прилога била је „физичка информација", назива који би требао истицати њену форму за коју важи закон одржања, наспрам Шенонове, тј. средње вредности Хатлијевих информација дате расподеле вероватноћа. Овде ћу кратко навести један од тамошњих многих примера са нагласком на „вишак" који крије „емергенцију".

Пример 1. (Бинарна расподела) Опит има жељени исход вероватноће $p \in (0,1)$, чиме је нежељени вероватноће $q = 1 - p$. Понављамо га $n = 1, 2, 3, \dots$ пута у непромењеним околностима и налазимо да се жељени исход појавио $k = 0, 1, \dots, n$ пута. То је модел типичне Бернулијеве расподеле $\mathfrak{B}(n, p)$, вероватноће

$$p_k = \binom{n}{k} p^k q^{n-k}$$

да ће се у n понављања жељени исход десити k пута. Биномни коефицијент је

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n \cdot (n-1) \cdots (n-k+1)}{1 \cdot 2 \cdots k}.$$

Очекивање и варијанса ове расподеле су $\mu = np$ и $\sigma^2 = npq$. За Шенонову информацију

$$S_n = -\sum_{k=0}^{n} p_k \log p_k,$$

не важи закон одржања ($S_n \ge nS_1$), уместо које можемо дефинисати „физичку информацију" са

$$L_n = -\sum_{k=0}^{n} p_k \log\left(p^k q^{n-k}\right) = -\sum_{k=1}^{n} \binom{n}{k} \left[p^k q^{n-k} \log\left(p^k q^{n-k}\right)\right]$$

за коју важи закон одржања ($L_n = nL_1$). При томе је $L_1 = S_1$ једина једнакост ове две. ❑

У случају мноштва, $n > 1$, рецимо удруживања, настаје „емергенција" $\Delta H_n = S_n - L_n$ која је растућа величина са порастом броја „удружених", тј. понављања n.

[17] Растко Вуковић: ФИЗИЧКА ИНФОРМАЦИЈА, Економски институт Бања Лука, 2019.

80. Условни догађаји

Да бисмо лакше разумели текст који следи замишљамо човека који куца слова неке азбуке, рецимо из скупа $\mathcal{A} = \{e_1, e_2, e_3, \dots\}$, али су примене много општије. То је дискретни извод података. Он куца насумице тако да производи низ независних случајних променљивих $x_1, x_2, x_3, \dots \in \mathcal{A}$ за који кажемо да има меморију нултог реда [64. Без памћења], или откуцано слово зависи најдаље од првог претходног када за добијени низ слова кажемо да има меморију првог реда. Уопште, ако је откуцано слово зависно до n-тог претходног кажемо да низ има меморију n-тог реда.

Рана истраживања у теорији информације показала су да говорни (енглески) језик има средњу меморију реда 8 – 10. Можемо погодити да у тексту „ТЕ.ЕФОН“ на месту тачке стоји слово „Л“ пре од слова нпр. „А“, ако је реч извађена из новина. Исто тако, реченица „БРОЈ ТЕЛЕФО“ настављаће са словима „НА“.

***

1. Оно што је интуитивно препознатљиво, у примерима сличним претходним, јесте да се неодређеност све дужег низа смањује. Када је сећање и присуство претходних образаца све веће, тада је фреквенција појаве слова у тексту све даља од експоненцијалне расподеле (Letter Frequency). Слично нешто имамо, на пример у гравитационом пољу. Вертикалан пад слабог, Њутновог поља који даје експоненцијалну зависност гравитационог потенцијала. Она у јачим пољима Ајнштајнових општих једначина, или истих изведених из принципа најмањег дејства (The Lagrangian), прелази у другачију врсту расподеле. Тада се дешава „повлачење простора“ за масом која орбитира виђено раније у случају померања Меркуровог перихела у смеру кретања планете око Сунца, а за које се такође (теорија информације) може рећи да је последица „памћења простора“ (AGC 114905).

2. Приметимо доследност ових појава у тумачењу случајности [27. Сила неизвесности]. Када исход мање зависи од прошлости он је више случајан, информативнији је, а крива расподеле таквих губи стрмост у односу на експоненцијалну. Недавно смо у расправи сличне теме наишли на још једно занимљиво питање. Колега се упитао зашто „радити више не значи зарадити више“, потакнут неким наводима из новина[18], тражећи одговор са становишта „теорије информације“.

Да, тачно је да особа која рутински ради, понављајући исти посао сваки дан, има мању плату од особе која доноси одлуке – одговарао сам – Када пажљивије истражујемо налазимо да начини рада који су ређи и због необичности бивају напорнији послови, знају бити боље плаћени, а још дубљим увидом дошли бисмо до препознавања неизвесности. Ето ту је „теорија информације“, управо у налазу „силе неизвесности“ као битног корена природних стања или покретача њених процеса, укључујући и стицање новца, моћи, угледа. Као да саме случајне појаве генеришу своје неслучајности. Условљавајући се догађаји се умрежавају, организују, постају ефикаснији, али на количини неизвесности, обиму опција – губе.

3. На слично условљавање своди се и само наизглед сасвим другачије питање из струке биолога о подсећању цивилизација на жива бића, по израстању младости, зрелости и старости. Да, нова и за сада будућа теорија информације даје дубљи смисао тим познатим поређењима – одговор је – Да више опција деловања има живо од неживог бића, већом количином неодређености влада, али и један и други немир теже смиривању. Попут настанка, развоја и престанка олује, живот је такође

[18] BBC, https://www.bbc.com/worklife/article/20170112-if-you-want-to-earn-more-work-less

„неприродна тежња“, противна принципу најмањег дејства, који попут камена баченог увис пре или касније пада савладан општом привлачном силом гравитације.

Кажем наводно (неприродна тежња), јер поремећаји су природна појава, као и различитости и уопште непредвидљивости. Штавише, иста одбојна „сила неизвесности“ која чини вероватније исходе чешћима, која преферира развоје у мање информативна стања, постаје равна са „силом различитости“ која би да бројност независних компоненти увећава.

Наиме, како систем (живо биће, цивилизација, титрање молекула течности) попушта у „количини неизвесности“ постајући усмеренији, губећи неизвесност (друштво се организује у „ефикасније“, а течност хладећи се кристалише), он шире гледајући због закона одржања неће губити своју укупну количину неизвесности, информацију. Систем тада утолико увећава своје независне компоненте колико се утеже и регулише. Другим речима, предефинисањем систем губи на виталности и нужно се распада.

Живот се у свом залету младости кад-тад нађе у зрелости и даље враћа мировању током старости. Најстарије клонско дрво на свету (норвешка смрека) старо је 9550 година и налази се на планини Фулуфјалет у шведској провинцији Далама. Открио га је професор физичке географије Кулман који га је назвао по свом псу старом „Тјику“. Верује се да је вегетативно клонирање и слојевитост узрок дуговечности ово дрвета. Током вегетативног клонирања, стабло дрвета може да умре на сваких стотинак година (око 600 у овом случају), али систем корења може да живи и даље и то хиљадама година. Слојевитост се дешава када једна од грана дотакне тло и временом „постане корен“. Век живота људи у просеку је мањи од 80 година, а неких инсеката тек је један дан.

Свођење на исто начело живих и неживих бића једна је од самих првих последица поставки (моје) теорије информације и један од разлога зашто не волим да је популаришем. Помало заобилазно, али једноставно писао сам о томе у „Причама“. Парафразирајући, већа ће међусобна условљеност делова целине снижавати њену укупну информацију и тражити допуну повећањем разноврсности структуре. Ако је поменута последица поставки ове теорије тачна, онда исто важи и за жива бића.

4. Више пута сам морао објашњавати да „живо“ биће од „неживог“ разликује вишак информације које оно поседује. Тај квантум чини границу виталности на вишем нивоу од кванта дејства, а са та два прага, нижег Планковог, који је дно познате квантне физике и, са друге стране, по сложености вишег од којег почиње биологија, два су од много делова „теорије информације“. Испод речених биле би неслободне информације које се физичким простором не могу кретати саме (ван кванта дејства), а оне далеке, углавном изнад ових, биле би бесконачности. Очекујемо да је условљеност сложенијих система све већа.

Спонтана је тежња природних процеса ка мање информативним (вероватнијим) стањима. Томе им стоји на путу закон одржања информације, а начини да прово начело (смањивања) заобиђе друго (конзервације) је путем многострукости[19] и памћења. Када разлика суседних „потенцијала“ (вероватноћа, информација) постаје превелика, као у прејаким гравитационим пољима, тада и простор почиње „да памти“. Ефекти тог „деловања прошлости на садашњост“ виде се у кретању

[19] Р. Вуковић: *МНОГОСТРУКОСТИ*, Економски институт Бања Лука, 2018. https://archive.org/details/Mnogostrukosti

елипсе путање Меркура, или у појавама које називамо „тамном материјом“. Заправо и у многим силама нашег свакодневног физичког окружења.

Укупна информација садашњости допуњена оном пристиглом нам из прошлости константна је. Међутим, васиона не постаје све развученија у времену само у том смислу, она је и све шира у временским паралелним димензијама, а уједно постаје и просторно све већа. Тела око нас зато су, релативно у односу на тренутке из прошлости, мало по мало инертнија (Growing), време тече све спорије (смањивање информације садашњости), брзина светлости све је мања (све даље галаксије чине нам се још све даље). Али допирања наше спознаје у дубине прошлости, ширину паралелних времена, такође и у просторне даљине имају своја ограничења [61.2. Ергодичка теорема]. Никада не можемо знати да је то што перципирамо (чулима, умовањем, уопште интеракцијама или научним знањима) све што „стварно“ постоји.

80.1. Низови

Полазећи од дефиниције условне вероватноће

$$\Pr(X_n|(X_1,\dots,X_{n-1}))=\frac{\Pr\,(X_1,\dots,X_{n-1},X_n)}{\Pr\,(X_1,\dots,X_{n-1})},$$

где је низ у загради пресек скупова случајних догађаја, добијамо

$$\Pr(X_1,\dots,X_{n-1},X_n)=\Pr(X_1,\dots,X_{n-1})\cdot\Pr\big(X_n\big|(X_1,\dots,X_{n-1})\big),$$

или због адитивности информације

$$S(X_1,\dots,X_{n-1},X_n)=S(X_1,\dots,X_{n-1})+S\big(X_n\big|(X_1,\dots,X_{n-1})\big).$$

Отуда рекурзијом следи

$$S(X_1,\dots,X_{n-1},X_n)=S(X_1)+S(X_2|X_1)+\dots+S\big(X_n\big|(X_1,\dots,X_{n-1})\big),$$

односно

$$S(X_1,\dots,X_{n-1})\le S(X_1,\dots,X_{n-1},X_n)\le S(X_1,\dots,X_{n-1})+S(X_n).$$

Лево важи једнакост ако и само ако је последњи скуп потпуно одређен претходним, а десно важи једнакост ако и само ако је последњи скуп независан од претходног. Интерпретирамо ли X_k као k-то слово у реченици, тада лева неједнакост значи да је дужа реченица више информативна, десна да слово изван текста носи више неодређености од слова у тексту.

Ове неједнакости говоре нам и о опадању информације када број слова текста расте, $n\to\infty$, при чему информација сваког следећег слова опада. Тај се закључак може формулисати прецизније

$$S\big(X_n\big|(X_1,\dots,X_{n-1})\big)\le S\big(X_n\big|(X_1,\dots,X_{n-2})\big).$$

Три последње релације се добијају и при разматрању канала. То проверавамо сменом

$$Y=(X_1,\dots,X_n),\quad X=X_{n+1},$$

па уместо средње добијамо $S(X) \leq S(X,Y) \leq S(X) + S(Y)$, а исто се добија са условним информацијама канала преноса $K: \vec{p} \to \vec{q}$. Обрнуто, из тих неједнакости канала након рачунања са коефицијентима могу се добити три послење релације, прво за дискретне расподеле, затим за континум.

Вероватноћу да је низ од n слова $e_1, \dots, e_n$ пишемо $\Pr\,(X_1 = e_1, \dots, X_n = e_n)$. Ако вероватноћа исечка од n слова не зависи од изабраног текста рећи ћемо да је текст стационаран. Прецизније, извор података је стационаран ако за свако $k, n = 1, 2, 3, \dots$ имамо

$$\Pr(X_{k+1} = e_1, \dots, X_{k+n} = e_n) = \Pr(X_1 = e_1, \dots, X_n = e_n).$$

Текст неке књиге може бити приближно стационаран ако теме које књига обрађује требају приближно константну количину појединих слова. Такви ће текстови имати приближно експоненцијалну расподелу, што значи да су и текстови чије сам фреквенције слова овде анализирао раније приближно такви. У наставку претпостављамо стационарност извора података.

Дефинишимо средње информације низа од n слова са:

$$s_n = S(X_n|X_1, \dots, X_{n-1}) = -\sum \Pr\,(e_1, \dots, e_n) \log \Pr(e_n|e_1, \dots, e_{n-1}),$$

$$S_n = \frac{1}{n} S(X_1, \dots, X_n) = -\frac{1}{n} \sum \Pr\,(e_1, \dots, e_n) \log \Pr(e_n|e_1, \dots, e_n),$$

где се сабира по свим комбинацијама n-чланих низова дате азбуке. Према томе је S_n средња својствена информација једног од n слова у низу, док је s_n информација коју даје n-то слово у низу када је познато претходних $n-1$.

Теорема 1.

i. Низ s_n је конвергентан.
ii. Низ S_n је конвергентан.
iii. $\lim_{n\to\infty} s_n = \lim_{n\to\infty} S_n = S < \infty$.

Доказ: Из горње неједнакости и стационарности следи

$$0 < s_n \leq S(X_n|X_1, \dots, X_{n-2}) = S(X_{n-1}|X_1, \dots, X_{n-2}) = s_{n-1}.$$

Дакле, низ s_n је ограничен и монотоно нерастући, па је он конвергентан, тј. $\lim_{n\to\infty} s_n = s < \infty$. Тиме је доказано (i).

Из горње збирне једнакости имамо:

$$S_n = \frac{1}{n}(s_1 + \dots + s_n) \geq s_n,$$

$$S_n = \frac{1}{n}[S(X_1, \dots, X_{n-1}) + S(X_n|X_1, \dots, X_{n-1})] = \frac{n-1}{n} S(X_{n-1}) + \frac{1}{n} s_n,$$

$$S_n \leq \frac{n-1}{n} S_{n-1} + \frac{1}{n} S_n,$$

односно $S_n \leq S_{n-1}$. Низ S_n је ограничен и монотоно нерастући, па је $\lim_{n\to\infty} S_n = S < \infty$. Тиме је доказано (ii).

Због $S_n > s_n$ $(n = 1, 2, \dots)$ је $S > s$. Са друге стране, због $s_n \leq s_{n+1}$ имамо

$$S_n = \frac{1}{m+n}[S(X_1, \dots, X_n) + s_{m+1} + \cdots + s_{m+n}] \leq \frac{1}{m+n}[S(X_1, \dots, X_n) + ns_{m+1}].$$

Пустимо ли да $n \to \infty$ добијамо $S \leq s_{m+1}$, а за $m \to \infty$ добијамо $S \leq s$, те $S = s < \infty$. ■

Информација $s = S$ у датом тексту представља средњу вредност информације коју садржи једно слово азбуке од a слова. Оне достижу максимум када се свако слово појављује једнак број пута у тексту и тада је $S = \log l$. У српском језику са $l = 30$ слова (у енглеском $l = 26$) плус размак, па је $S_{\text{srb}} = \log_2 31 \approx 4{,}95$ ($S_{\text{eng}} = \log_2 27 \approx 4{,}75$), док је информација говорног[20] српског језика 4,31 (енг. 4,14). Што је разлика између ове две, максималне и стварне информације слова дате азбуке, већа – то је текст садржајнији, смисленији. Такође, ово смањење информације, као и редунданса $(1 - S/\log l)$, помаже да одгонетнемо нејасне, погрешне или скраћене речи.

80.2. Понављања

Опет ћу рећи нешто о законитости великих бројева, сада са становишта редундансе (понављања речи истог значења), односно условљених догађаја. Попут боље организованог система, доследан текст који поседује предвидљост, или обрнуто речено који не поседује сувишне опције – мање је информативан. Он наликује преопширном излагању са пуно сувишних, непотребних речи, или понављања саопштења у комуникацији како би се избегао неспоразум. Тај вишак функционално истих елемената у једном механизму ради појачане сигурности његовог деловања, који га чини редундантним, који му смањује неизвесност, повећава му поузданост.

Овај „формализам редундансе“ можемо препознати у физици мноштва честица термодинамике, који налаже да се молекуле гаса у соби распоређују на случајан начин, али тако да се поравнавају у средње вредности (математичка очекивања). Оне се зато никада неће све затећи у једном ћошку собе остављајући нас да умремо у вакууму у делу без ваздуха. Слично просторном распоређивању „понављања“, дешава се временско распоређивање развојем васионе у све извеснију, све дужег памћења о којем сам више пута писао, а сада са нагласком на развој око једне средње вредности.

У (хипо)тези да се време садашњости све више шири у паралелна времена (универзуме), такође је присутно ово усредњавање. Са порастом извесности наше садашњости њена информација све је мања, њена разуђеност све је већа и дебљина на случајан јој начин доступних временских псеудо реалности увећава се. При томе, свака од паралелних реалности ограничена је слично нашој, она поседује много веће ширине од из ње „оптички видљивих“, а токови сваке од њих као да све више личе једни другима. То је попут континуума правих из исте тачке које суседне, а удаљавањем од заједничког исходишта, све све више личе на паралелне.

У хипотетичкој теорији према којој је информација ткање простора, времена и материје, а њена суштина је неизвесност, у таквој је и сама васиона неизвесност у оба смера, изнутра и извана. Та очекивања потврђује и тополошки приступ овој проблематици (Dimensions) из којег следи једнак

[20] Letter Frequency, https://www.academia.edu/86260704/Letter_Frequency

број временских и просторних димензија, односно бар 6 димензионалност простор-времена наше васионе. Затим просто ширимо перспективу на одговарајуће 4-дим њене делове.

***

Вратимо се сада на Чебишевљеву неједнакост коју смо и овде [76.2.] доказивали

$$\Pr(|X| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Полазећи од произвољног низа $X_1, X_2, \dots, X_n$, у којем случајна променљива X_k има математичко очекивање μ_k, дефинишемо нову променљиву

$$Y_n = \frac{1}{n}\sum_{k=1}^{n}(X_k - \mu_k).$$

Очекивање ове променљиве је:

$$E(Y_n) = \frac{1}{n}\sum_{k=1}^{n}(E(X_k) - \mu_k) = 0,$$

па је варијанса:

$$\sigma^2(Y_n) = E(Y_n - E(Y_n))^2 = \frac{1}{n^2}\sigma^2\left(\sum_k X_k\right).$$

Непосредно из Чебишевљеве неједнакости сада следи следећи тзв. Марковљев закон.

Ако је низ случајних променљивих $X_1, X_2, \dots$ такав да

$$\frac{1}{n^2}\sigma^2\left(\sum_k X_k\right) \to 0$$

када $n \to \infty$, тада за сваки $\varepsilon > 0$

$$\lim_{n\to\infty} \Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n}\mu_k\right| \geq \varepsilon\right) = 0.$$

Посебно, ако су променљиве независне, због

$$\sigma^2\left(\sum_k X_k\right) = \sum_k \sigma^2(X_k)$$

имамо познати Чебишевљев закон великих бројева [26. Пример 2]. Према томе је Марковљев закон поопштење Чебишевљевог за низ зависних случајних променљивих.

***

Интуитивни смисао Марковљевог закона великих бројева јаснији је ако претпоставимо да је свако $\mu_k = 0$. Тада

$$\Pr\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right| \geq \varepsilon\right) \to 0$$

када $n \to \infty$. Дакле, аритметичка средина низа бројева $X_1, X_2, \ldots$ се групише око очекивања једног X_k $(k = 1, 2, \ldots)$. Поменути закон даје потребан, али не и довољан услов за такво груписање. Ради тога следећи став.

Теорема 2. За произољан низ случајних бројева $Y_1, Y_2, \ldots$ вреди $\Pr(|Y_n| \geq \varepsilon) \to 0$, када $n \to \infty$, за свако $\varepsilon > 0$, ако и само ако очекивање

$$E\left(\frac{Y_n^2}{1 + Y_n^2}\right) \to 0$$

када $n \to \infty$.

Доказ: Према Чебишевљевој неједнакости имамо:

$$\Pr(|Y_n| \geq \varepsilon) \leq \frac{\sigma_n^2(Y_n)}{\varepsilon^2} = \frac{1+\varepsilon^2}{\varepsilon^2} \cdot \frac{\sigma^2(Y_n)}{1+\varepsilon^2} \leq \frac{1+\varepsilon^2}{\varepsilon^2} \cdot E\left(\frac{Y_n^2}{1+Y_n^2}\right) \to 0.$$

Тиме је доказан први део теореме. Претпоставимо да је Y_n непрекидна промењљива. Тада је:

$$\Pr(|Y_n| \geq \varepsilon) = \int_{|y| \geq \varepsilon} dF_n(y) \geq \int_{|y| \geq \varepsilon} \frac{y^2}{1+y^2} dF_n(y) =$$

$$= \int \frac{y^2}{1+y^2} dF_n(y) - \int_{|y| < \varepsilon} \frac{y^2}{1+y^2} dF_n(y) = E\left(\frac{Y_n^2}{1+Y_n^2}\right) - \varepsilon^2.$$

Сличну неједнакост добијамо за дискретно Y_n. Када $\varepsilon \to 0$, а затим $n \to \infty$, доказујемо други део теореме. ■

Очекивање и варијанса су линеарни оператори, па уместо $E(X_n)$ и $\sigma^2(X_n)$ можемо писати краће EX_n и $\sigma^2 X_n$. Због:

$$\frac{Y_n^2}{1+Y_n^2} \leq Y_n^2 = \left[\frac{1}{n}\sum_{k=1}^{n}(X_k - EX_k)\right]^2 = \frac{1}{n^2} \cdot \sigma^2\left(\sum_{k=1}^{n} X_k\right)$$

последица ове теореме је Марковљев закон великих бројева. Наизглед безазлено и очигледно тврђење 2. теореме у себи крије и Бернулијев закон великих бројева, што ћемо видети у следећем примеру.

Пример 1. Статистичка вероватноћа (p) неког догађаја $(X = e)$ се дефинише као однос броја појављивања (x) тога догађаја и укупног броја (N) поновљених независних опита. Догађај се у једном опиту или десио (са вероватноћом p) или се није десио $(q = 1 - p)$. Тада је $EX = p$, $\sigma^2 X = pq \leq \frac{1}{4}$, па из Чебишевљеве неједнакости непосредно добијамо да за свако $\varepsilon > 0$ је

$$\Pr\left(\left|\frac{x}{N} - p\right| \geq \varepsilon\right) \to 0$$

када $N \to \infty$. То је Бернулијев закон великих бројева за независне догађаје, који је специјални случај 2. теореме, за $Y_n = \frac{x}{N} - p$. ❑

Пример 2. Када имамо низ од $N = 1, 2, \dots$ независних опита у којем се k-ти догађај, вероватноће p_k, реализовао $X = 0, 1, \ \dots, N$ пута, тада важи Пуасонов закон великих бројева

$$\lim_{N\to\infty} \Pr\left(\left|\frac{X}{N} - \frac{p_1 + p_2 + \cdots + p_N}{N}\right| \geq \varepsilon\right) = 0,$$

за свако $\varepsilon > 0$. Наиме, у k-том опиту је $EX_k = p_k$, $\sigma^2 X_k = p_k \cdot q_k \leq \frac{1}{4}$, где је $q_k = 1 - p_k$, па због $EX_1 + \cdots + EX_N = EX = p_1 + \cdots + p_N$ и независности опита $\sigma^2 X_1 + \cdots + \sigma^2 X_N = \sigma^2 X = N \cdot \sigma^2 X_1$, те је:

$$\Pr\left(|X - EX| \geq \varepsilon\right) \leq \frac{\sigma^2 X}{N^2 \cdot \varepsilon^2} \to 0$$

када $N \to \infty$. ❑

У оба наведена примера низ опита је независан, а низ $X_1, X_2, \dots$ ергодички.

Подсећам, у физици, статистици, економетрији и обради сигнала, за стохастички процес се каже да је у ергодичком режиму ако је просек опсервабилног ансамбла једнак временском просеку. У овом режиму, свака колекција насумичних узорака из процеса мора представљати просечна статистичка својства целог режима. Насупрот томе, за процес који није у ергодичком режиму каже се да је у неергодичком режиму.

Замислимо, међутим, исти низ опита као у горњим примерима, али тако да је X_k укупан број реализација до закључно k-тог опита. Тада низ X_k није независан, јер очито зависи од збира свих претходних реализација. Ако се дати догађај уопште дешава, тј. ако је његова вероватноћа позитиван број, онда очекујемо да случајна промењљива X_k неограничено расте када $k \to \infty$. Тада имамо једну дивергентну расподелу са дисперзијом која не задовољава услов Марковљевог закона.

Када низ $Y_1, Y_2, \dots$ има растуће дисперзије, тада су и дисперзије низа $X_1, X_2, \dots$ растуће, па за њих не важи закон великих бројева. Тада нема нагомилавања вредности низова у граничном случају, односно они немају граничне вероватноће. Обрнуто, ако дати низови немају граничне вероватноће, то значи да им вредности за $n \to \infty$ „осцилирају“, што другим речима значи да им дисперзије не теже нули.

81. Ергодички извор

Својство ергодичности у суштини значи могућност преласка стања у било које друго стање. Тако се ергодичка теорема [61.2] односи на матрице преноса информације које имају макар колико мале, али све врсте грешака. Почетни сигнал постепено, преносећи се дугим низом таквих матрица, све више постаје својствени вектор дате матрице генератора канала, а канал трансформације на крају је „црна кутија“. Свака улазна порука адаптира се у исту излазну.

У стварности, на пример еволуцијом живих бића на земљи, преноси се генетска информација кроз окружење, али тако да се обе прилагођавају једна другој. Важи ергодичност иако немамо један те исти, константан канал преноса, па почетно окружење заједно са живим бићима унутар, процесом еволуције све мање личи каснијем. Слична овој је и раније помињана „еволуција“ васионе.

Пример 1. Нека за вероватноће важи $a+b=1$ и $p+q=1$, где $a,b,p,q>0$. Оне су коефицијенти матрице ергодичког канала

$$\hat{K}=\begin{pmatrix} a & b \\ p & q \end{pmatrix}.$$

Својствена једначина $\hat{K}\vec{x}=\lambda\vec{x}$, даје:

$$\begin{pmatrix} a & b \\ p & q \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}=\lambda\begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

$$\begin{cases} (a-\lambda)x_1+bx_2=0, \\ px_1+(q-\lambda)x_2=0. \end{cases}$$

Да би овај хомогени систем једначина имао нетривијално решење (поред $x_1=x_2=0$) мора му детерминанта бити нула:

$$\begin{vmatrix} a-\lambda & b \\ p & q-\lambda \end{vmatrix}=0,$$

$$\lambda^2-(b+q)\lambda+bp-aq=0,$$

што је увек тачно за својствену вредност $\lambda=1$. Тада из горњег система добијамо $x_1=x_2=1/2$, где је урачунато да је $\vec{x}$ вектор расподеле. ❑

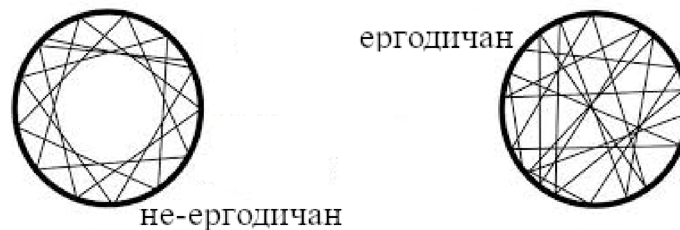
Својствена бинарна расподела са две исте вероватноће, матрице овог примера, резултат је сваке поруке преношене кроз канал дугог низа њоме генерисаног. Та коначна порука не дефинише ону полазну нити мало, а од свих могућих бинарних вектора садржи максималну информацију. Рекли смо да је максимална информација излаза таква јер се улазној шумом информација додаје (иако је уобичајено рећи да се она тако „губи“).

Супротна овој композитној последици ергодичности је расподела по различитим вредностима тзв. AEP (Asymptotic Equipartition Property) у локалу, коју имају сви ергодички извори. Ергодички извор се односи на процес у којем су сви низови или њихови велики узорци, једнако репрезентативни за целину. Такође, појам ергоичког извора укључује или се односи на вероватноћу да ће се било које стање поновити са нултом вероватноћом да се било које стање никад не понови. Природа воли да свакоме да неку шансу, али не воли једнакост (Equality).

Претпостављамо и даље да је говорни језик стационаран и ергодички (неких шанси за све), тј. да је низ изговорених слова стационаран и ергодички низ, што ћемо касније искористити за тражење оптималног кода такве класе језика, затим модел применити и протумачити у другим ситуацијама.

Тако, листајући опширан речник може нас изненадити колико је много ретко кориштених израза у свакодневном животу. Без обзира што су они свима нама познати, прислушкивајући разговоре на улици по цео дан безуспешно ћемо чекати да их неко употреби. Ову особину ретких речи обично правдамо непотребношћу, ризиком да нас саговорник неће разумети, или на трећи начин, често субјективан, али стварни разлог може бити далеко од тога.

Показује се да је поменути опис особина ергодичког језика, који се може поделити на скуп ниско и високо вероватних речи, рецимо A_0 и A_1, што и није неко изненађење, али јесте да вероватноћа првог тежи нули ($\Pr\ A_0 \to 0$) и да је он (A_0) велики подскуп језика. Другим речима, давање шансе свима рађа законитост да велики део њих чини безначајнима!



Пример 2. Бацамо новчић 100 пута[21] и добијамо nиз падања „главе“ (Head) и „писма“ (Tail), H, H, H, H, H, T, ..., H, H, H, T, T, у којем се прво јавља 58 пута а друго 42. Када бисмо поновили радњу, не би добили исти низ H-T, нити можда исти однос 58 : 42, иако су шансе ова два исхода пола-пола. У неком трећем, четвртом, ..., n-том понављању оваквих 100 бацања новчића добијали бисмо разне од ових $100! = 1 \cdot 2 \cdot 3 \cdots 100$ начина падања, али неки би се појављивали ређе од других. Оних најређих било би највише, а шанса да се појаве била би ништавна. ❑

Даље поопштење овог примера са новчићем, са два једнако вероватна исхода, је биномна расподела [34.] са n понављања (у примеру $n = 100$) са вероватнћом повољног исхода $p \in (0,1)$ и неповољног (свих осталих) $q = 1 - p$, константних позитивних бројева. У примеру $p = q = 0{,}5$. Слике, графици примера биномне расподеле у поменутом наслову показују како се исходи групишу када тражимо $k = 0, 1, \dots, n$ повољних исхода у n понављања, са вероватноћом

$$p_k(n) = \binom{n}{k} p^k q^{n-k}.$$

Овај број је максималан око очекивања ($\mu = np$), са брзим опадањем около.

81.1. Раздвајање

Знамо да је $I = -\log \Pr(e)$ информација слова e неке азбуке $\mathcal{A}$, где је $\Pr(e)$ вероватноћа појављивања тог слова. Да бисмо дошли до информације I у дугачком тексту можемо поћи од реченице дужине $n = 1,2,3, \dots$ слова и средње информације по слову

[21] Flip a Coin 100 Times, https://flip-a-coin-tosser.com/coin-toss-100-times/

$$I_n(x_1, \dots, x_n) = -\frac{1}{n} \log \Pr(x_1, \dots, x_n),$$

где је $\Pr(x_1, \dots, x_n) = p$ вероватноћа поруке $\vec{x} = (x_1, \dots, x_n)$. Како је број p случајан, то је $I_1, I_2, I_3, \dots$ случајан низ. Математичко очекивање таквог низа је:

$$EI_n = -\sum \Pr(\vec{x}) \log \Pr(\vec{x}) = \frac{1}{n} S(\vec{x}) = S_n$$

где се сабира по свим n-чланим низовима $\vec{x}$ азбуке $\mathcal{A}$. Према теореми [80.1. Низови, 1] је

$$\lim_{n\to\infty} EI_n = S.$$

Познајући лимес очекивања ми још увек не знамо $I = \lim_{n\to\infty} I_n$.

Друго, нисмо поделили језик на ниско и високо вероватне речи. Ради тога приметимо да се очекивање може поделити на три дела

$$EI_n = \sum_{I_n < \alpha - \gamma} + \sum_{\alpha - \gamma \le I_n \le \alpha + \beta} + \sum_{I_n > \alpha + \beta},$$

где су $\alpha \ge 0$, $\beta > 0$ и $\gamma > 0$ произвољни. За први од ових сабирка је очигледно

$$\sum_{I_n < \alpha - \gamma} < (\alpha - \gamma) \sum_{I_n < \alpha - \gamma} P_n = (\alpha - \gamma) \Pr(I_n < \alpha - \gamma),$$

а за другу и трећу партицију је:

$$\sum_{\alpha - \beta \le I_n \le \alpha + \beta} \le (\alpha + \beta) \Pr(\alpha - \gamma \le I_n \le \alpha + \beta) \le$$

$$\le (\alpha + \gamma) \Pr(I_n \ge \alpha - \gamma) = (\alpha + \gamma)[1 - \Pr(I_n < \alpha - \beta)],$$

$$\sum_{I_n > \alpha + \beta} \le \Pr(I_n > \alpha + \beta) \log \Pr(I_n > \alpha + \beta).$$

Нека је l број слова азбуке, а $A \subseteq \mathcal{A}$ подскуп језика састављен управо од речи чија нам вероватноћа треба, тј.

$$A = \{(x_1, \dots, x_n) :\ I_n > \alpha + \beta\}.$$

Тада је $\sum \Pr(\vec{x}) = \Pr(A) \le 1$, $\Pr(\vec{x}|A) = 1$ где се сабира по свим низовима $\vec{x} \in A$. Сада је

$$S(A) \le n \cdot \log l,$$

јер је на десној страни неједнакости максимална информација n-чланих низова (α различитих елемената). На левој страни је:

$$S(A) = -\sum_{\vec{x} \in A} \Pr(\vec{x}|A) \log \Pr(\vec{x}|A) = \sum_{\vec{x} \in A} \frac{\Pr(\vec{x})}{\Pr(A)} \log \frac{\Pr(\vec{x})}{\Pr(A)},$$

те је

$$-\frac{1}{n}\sum_{\vec{x}\in A}\Pr(\vec{x})\log\Pr(\vec{x}) \le \Pr(A)\log a - \frac{1}{n}\Pr(A)\log\Pr(A).$$

То је релација за трећи од горњих сабирака. Сва три сабирка заједно дају

$$EI_n - \alpha \le \beta - (\beta+\gamma)\Pr(I_n < \alpha - \gamma) +$$

$$+\Pr(I_n > \alpha + \beta)\log a - \frac{1}{n}\Pr(I_n > \alpha+\beta)\log\Pr(I_n > \alpha+\beta).$$

Последњу релацију и њен гранични облик користићемо за доказ наредна два става.

81.2. Две теореме

Нека је:

$A_1^{(n)}$ – високовероватан подскуп језика, са l слова, састављен од речи дужине n;

L – максимално I_n за фиксно n језика, тј. $L = \frac{1}{n}\log_b l$ или $l = b^{nL}$;

$Q_n(L)$ – вероватноћа високо вероватног скупа $A_1^{(n)}$, тј.

$$Q_n(L) = \Pr\left(A_1^{(n)}\right) = \max_{x_i\in\mathcal{A}} \Pr(\vec{x}),\ \vec{x} = (x_1, \dots, x_n).$$

Теорема 1. Важе тврђења:

i. $Q_n(L) \le \Pr(I_n \le L + \beta) + b^{-n\beta}$, $\beta > 0$, $b > 1$;
ii. Ако за свако $\beta > 0$ је $\lim_{n\to\infty}\Pr(I_n > S + \beta) = 0$, тада је $\lim_{n\to\infty}\Pr(|I_n - S| \le \beta) = 1$;
iii. Ако $\forall\beta > 0$ постоји природан број m_0 такав да за сваки $m > m_0$ је $\lim_{n\to\infty} Q_n(h_m + \beta) = 1$, где је h_m дефинисано [80.1. Теорема 1], па $(\forall\beta > 0)\ \lim_{n\to\infty}\Pr(|I_n - S| \le \beta) = 1$.

Доказ: Раставимо скуп свих речи дужине n на два дисјунктна подскупа:

$$C_1 = \{\vec{x} :\ I_n \le L + \beta\},\quad C_2 = \{\vec{x} :\ I_n > L + \beta\}.$$

Тада је:

$$Q_n(L) = \Pr\left(A_1^{(n)}\right) = \Pr\left(A_1^{(n)} \cap C_1\right) + \Pr\left(A_1^{(n)} \cap C_2\right) \le \Pr(C_1) + \Pr\left(A_1^{(n)} \cap C_2\right).$$

За $\vec{x} \in A_1^{(n)} \cap C_2$ је $\Pr(\vec{x}) < e^{-n(L+\beta)}$, те је $\Pr\left(A_1^{(n)} \cap C_2\right) < le^{-(L+\beta)} = e^{-n\beta}$, јер скуп $A_1^{(n)} \cap C_2$ има мање елемената од скупа $A_1^{(n)}$ а овај их нема више од l. Сменом у претходну једнакост добијамо доказ прве ставке (i).

Доказ за (ii) добијамо из претходног лимеса и доказа треће партиције. Сменом $\beta = S = EI_n$, биће

$$\lim_{n\to\infty}\sup\Pr(I_n < EI_n - \gamma) \le \frac{\beta}{\beta+\gamma}.$$

Како је по претпоставци $\beta > 0$ произвољно, то је последњи лимес нула. Стављајући $\gamma = \beta$ излази доказ за (ii).

На основу (i) стављајући $L = h_m + \beta$ добијамо

$$\Pr(I_n < h_m + 2\beta) \geq Q_n(h_m + \beta) - b^{-n\beta},$$

те је

$$\lim_{n\to\infty} \Pr(I_n \leq h_m + 2\beta) = 1, \quad m > m_0.$$

Како је $h_m \to S$ нерастући низ то постоји $\beta > 0$ такво да је $h_m \leq S + \frac{\beta}{2}$ или $h_m + \frac{\beta}{2} < S + \beta$. Отуда

$$\Pr\,(I_n < S + \beta) \geq \Pr\,(I_n \leq h_m + \frac{\beta}{2}) \to 1,$$

$$\Pr\,(I_n < S + \beta) \to 1.$$

Због (ii) добијамо (iii). Тиме је теорема доказана. ■

У управо доказаној теореми (недоказане) претпоставке (ii) и (iii) дају последицу $Q_n \to 1$, $I_n \to S$. Са следећим доказом тих претпоставки потврдићемо постојање подскупа језика вероватноће и доказати да је информација по слову једнака информацији језика. Ова последња етапа је могућа за ергодичке изворе.

Теорема 2. Нека је $S > 0$ информација стационарног ергодичког извора и

$$I_n = I_n(\vec{y}) = -\frac{1}{n}\log_b \Pr\,(\vec{y})$$

информација по слову у речи дужине $n = 1,2,3,\ldots$ од $l = 1,2,\ldots,n$ различитих слова. Тада је

$$\lim_{n\to\infty} \Pr\,(|I_n - S| \geq \varepsilon) = 0$$

за свако $\varepsilon > 0$.

Доказ: Нека је дата реч $\vec{x} = (x_1, x_2, \ldots, x_m)$, $x_i \in \mathcal{A}$, $m \leq n$. При емитовању низа од n слова та реч се појавила v_n пута. Понављајући поступак долазимо до статистичке вероватноће v_n/n појаве речи $\vec{x}$. Претпоставка да је извор ергодичан значи да постоји број $p = \Pr\,(\vec{x})$ такав да за $n \to \infty$ буде $v_n \to p$. Дефинишимо S_n скуп свих низова дужине n у којем се дата реч $\vec{x}$ појављује са учесталошћу v_n/n различитом од p за мање од $\beta > 0$. Прецизније

$$C_n = \left\{\vec{x} = (x_1, x_2, \ldots, x_m), x_i \in \mathcal{A}, m \leq n \,:\, \left|\frac{v_n}{n} - \Pr\,(\vec{x})\right| \leq \beta\right\}.$$

Показаћемо да је

$$\lim_{n\to\infty} \Pr\,(\vec{x}) = 1.$$

Комплемент скупа C_n је

$$\bar{C_n} = \bigcup_{\vec{x}} \left\{\vec{x} \,:\, \left|\frac{v_n}{n} - \Pr\,(\vec{x})\right| > \beta\right\},$$

те је:

$$\Pr(\bar{C}_n) = 1 - \Pr(C_n) \leq \sum_{\vec{x}} \Pr\left(\left|\frac{v_n}{n} - \Pr(\vec{x})\right| > \beta\right) < \sum_{\vec{x}} \varepsilon$$

због ергодичности извора, чим је $n \geq n_0$. Како свих речи $\vec{x}$ дужине n има мање од l^n то десни збир не прелази $l^n\varepsilon$. Према томе је

$$\Pr(C_n) > 1 - l^n\varepsilon.$$

Како је $\varepsilon > 0$ произвољно, а n фиксирано, то је тачан горњи лимес.

Показаћемо да у скупу C_n нема више од $b^{n(h_m+\beta)}$ елемената, где је $b > 1$ база логаритма информације. За произвољан $\vec{y} = (y_1, y_2, \dots, y_n)$ дефинишимо функцију

$$f(\vec{y}) = f(y_1, \dots, y_m, y_{m+1}, \dots, y_n) = \Pr(y_1, \dots, y_{m-1}) \cdot \prod_{i=m}^{n} \Pr\big(y_i \big| (y_{i-m+1}, \dots, y_{i-1})\big).$$

По дефиницији, за $\vec{x} \in C_n$ је:

$$\frac{v_n}{n} \leq \Pr(\vec{x}) + \beta,$$

$$v_n < n \cdot [\Pr(\vec{x}) + \beta].$$

Отуда

$$f(\vec{y}) = \Pr(y_1, \dots, y_{m-1}) \cdot \prod_{\vec{x}} \left[\Pr\big(x_m \big| (x_1, \dots, x_{m-1})\big)\right]^{v_n},$$

јер међу условним вероватноћама има v_n једнаких, па је:

$$f(\vec{y}) \geq \Pr(y_1, \dots, y_{m-1}) \cdot \prod_{\vec{x}} \left[\Pr\big(x_m \big| (x_1, \dots, x_{m-1})\big)\right]^{n \cdot [\Pr(\vec{x})+\beta]},$$

$$[f(\vec{y})]^{1/n} \geq [\Pr(y_1, \dots, y_{m-1})]^{\frac{1}{n}} \cdot \prod_{\vec{x}} \left[\Pr\big(x_m \big| (x_1, \dots, x_{m-1})\big)\right]^{\Pr(\vec{x})+\beta}.$$

Због ергодичности извора за свако $\varepsilon > 0$ постоји природан број n_0 и такво $\beta_0 > 0$ да за све $n \geq n_0$ и $0 < \beta < \beta_0$ важи

$$[\Pr(y_1, \dots, y_{m-1})]^{\frac{1}{n}} \cdot \prod_{\vec{x}} \left[\Pr\big(x_m \big| (x_1, \dots, x_{m-1})\big)\right]^{\beta} \geq b^{-\varepsilon}.$$

Тако је

$$[f(\vec{y})]^{1/n} > b^{-\varepsilon} \cdot \prod_{\vec{x}} \left[\Pr\big(x_m \big| (x_1, \dots, x_{m-1})\big)\right]^{\Pr(\vec{x})}.$$

Логаритмовање по бази $b > 1$ и због дефиниције h_m излази:

$$\log_b f(\vec{y}) \geq n \cdot (-\varepsilon - h_m),$$

$$f(\vec{y}) \geq b^{-n \cdot \cdot (h_m+\varepsilon)}.$$

Означимо са N_n број елемената скупа C_n. Због:

$$N_n \cdot b^{-n(h_m+\varepsilon)} \leq \sum_{\vec{x} \in C_n} f(\vec{x}) \leq 1,$$

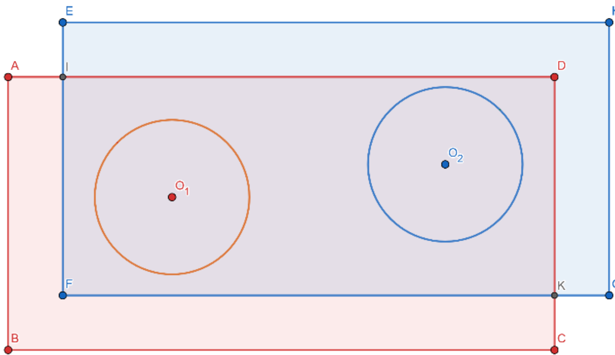
$$N_n \leq b^{n(h_m+\varepsilon)},$$

а то је управо горње тврђење о броју елемената скупа C_n. Са друге стране, доказали смо и предуслов за помоћни став. ■

Ове дугачке доказе тешких теорема многи неће читати, а и од оних који би можда то хтели многи не би разумели. Поједностављено речено, имамо низ $\omega_1, \omega_2, \dots, \omega_N$ независних догађаја, делове неког скупа Ω. Њихове вероватноће $p_i = \Pr(\omega_i)$ чине расподелу (свако p_i веће је од 0 и збир свих је 1), са средњом вредношћу информације $S = -p_1 \log p_1 - p_N \log p_N$. Највећу такву информацију ($\max S = \log N$) имаће скуп једнако вероватних исхода ($p_i =$ const. $= 1/N$). Али, тада свако $p_i \to 0$, када $N \to \infty$, па је тврђење теорема очигледно тачно. Такође је очигледно тачно тврђење теорема ако је неки (мањи) део свих исхода значајно вероватнији од осталог (већег) дела.

Оно што је теже запазити у интерпретацијама је да важи начело штедње емисије информације, да су вероватнији исходи чешћи догађаји а такви су мање информативни, те ће се стања развијати ка различитим вероватноћама. То је могуће уз одржање укупне информације рецимо повећањем N броја независних исхода. Друго, у случају ергодичности (неких шанси свих), реалност дела исхода подразумева псеудо-реалност још већег дела. Пример првог биле би тешкоће утезања глобализације са „побуњеним групама“, због једнаке силе која води у распадање целине. Пример другог је огроман видљиви део свемира, што значи да је светлошћу нама невидљиви део још већи.

Због произвољности разлике ($\forall \beta > 0$) део скоро нулте вероватноће много је већи од оног који се какав такав реализује (теорема 1). То са становишта неког посматрача (из центра O_1 круга на слици лево) њему репрезентативног дела миноран део може имати другог посматрача (из центра O_2 десног круга те слике) коме ће управо такви бити важни, али са такође пуно више делова њему неважних. Стога значајан део изван важности оба ова два посматрача ($O_1 \cup O_2$) лежи „тамо негде“. Другим речима, прихватање претпоставке ергодичности (те и ове теорије информације) слаже се са „немогућности скупа свих скупова“ (Раселов парадокс), или са „ограничењем поседовања знања о свему“ (Геделова теорема немогућности).



82. Асимптотска подела

Асимптотска подела за општи ергодички извор следи из претходних теорема. Докази су били прилично тешки и компликовани па је бит проблема остала нејасна. Укратко, имамо следеће. Низ случајних променљивих $x_1, x_2, \ldots$ нека је у општем случају бесконачан, стационаран и ергодички. Све променљиве могу узимати било коју од вредности из коначног скупа слова $\mathcal{A} = \{e_1, e_2, \ldots, e_l\}$. То је азбука, чији произвољан подниз називамо реч или реченица. Скуп свих речи је језик $\mathcal{J}$.

Уводна теорема [80.1. Низови] тврди да сваки стационарни извор са коначном азбуком има коначну информацију, а ако је уз то извор ергодичан онда теореме [81.2. Две теореме] тврде да је информација извора једнака просечној информацији једног елемента низа (x_n), тј. слова у тексту. При томе стационарност значи да се произвољно слово e_i у сваком поднизу појављује са истом вероватноћом p_i, а ергодичност значи да иста особина остаје и за бесконачне поднизове укључујући и цео низ $x_1, x_2, \ldots$.

Уз последње теореме такође је доказано да се речи стационарног, ергодичког језика $\mathcal{J}$ могу раздвојити у два подскупа:

A_0 – ниско вероватних речи,

A_1 – високо вероватних речи,

тако да је:

(1) $A_0 \cap A_1 = \emptyset, \qquad A_0 \cup A_1 = \mathcal{J};$

(2) $\Pr(A_0) = 0, \qquad \Pr(A_1) = 1;$

(3) $\text{No}(A_1, n) = 2^{n \cdot S}, \quad \text{No}(\mathcal{J}, n) = 2^{-n \cdot \log_2 l}.$

У трећој једнакости скраћеница „No“ значи „број“, а $\text{No}(A, n)$ је број речи дужине n у скупу A, база логаритма је 2 и информација је извора изражена у битима. Када су бар неке вероватноће слова p_i међусобно различите, тада $S < \log_2 l$, па број

$$\frac{\text{No}(A_1, n)}{\text{No}(\mathcal{J}, n)} = 2^{n(S - \log_2 l)}$$

(брзо) тежи нули када $n \to \infty$. У том је случају број високовероватних речи занемарљив део свих речи језика. Подела језика (1), (2), (3) је асимптотска ергодичка подела (АЕП).

Ове особине је могуће и доказивати и појединачно, на посебним случајевима извора, што ћемо демонстрирати у следећа три примера наведена редом према тежини.

82.1. Бернулијев случај

Нека слова $x_1, x_2, x_3, \ldots$ извора узимају само две вредности, тј. $\mathcal{A} = \{e_1, e_2\}$, са вероватноћама

$$\Pr(x_i = e_1) = p, \quad \Pr(x_i = e_2) = q, \quad i = 1, 2, 3, \ldots$$

То је стационарни извор са бројем слова $l = 2$, какав смо већ третирали [34. Бинарне расподеле] са другачијим циљевима, ергодички је са информацијом

$$S = -p \cdot \log_2 p - q \cdot \log_2 q \leq 1,$$

где једнакост важи ако и само ако $p = q$. Ово је и средња информација слова, тзв. Шенонова.

Реч велике дужине n садржи приближно np слова e_1 и nq слова e_2. Број речи је приближно

$$N_n = \frac{n!}{(np)! \cdot (nq)!}.$$

Према Стирлинговој формули имамо:

$$\log_2 N_n = \log_2 \frac{\sqrt{2\pi n} \cdot n^n \cdot e^{-n}}{\left[\sqrt{2\pi np} \cdot (np)^{np} \cdot e^{-np}\right] \cdot \left[\sqrt{2\pi nq} \cdot (nq)^{nq} \cdot e^{-nq}\right]} =$$

$$= n \log_2 n - np \log_2 np - nq \log_2 nq + \left(\log_2 \sqrt{2\pi n} - \log_2 \sqrt{2\pi np} - \log_2 \sqrt{2\pi nq}\right)$$

$$= n(-p \log_2 p - q \log_2 q) = nS,$$

или $N_n = 2^{nS}$, у бинарној бази логаритама.

Дакле, речи које садрже np слова e_1 и nq слова e_2 у скупу свих речи дужине $np + nq = n$ има приближно 2^{nS}, а тај резултат је утолико тачнији што је број n већи. Иначе је број свих бинарних (двословних) речи дужине n тачно 2^n ($n = 1, 2, 3, \dots$).

Пример 1. При бацању коцке дефинишемо слова $\mathcal{A} = \{e_1, e_2\}$:

e_1 – пала је шестица;

e_2 – није пала шестица.

Овде имамо низ бацања коцке са резултатима $x_1, x_2, x_3, \dots$ где је у i-том бацању:

$$\Pr(x_i = e_1) = \frac{1}{6}, \quad \Pr(x_i = e_2) = \frac{5}{6}, \quad i = 1, 2, 3, \dots$$

што је један Бернулијев извор података. Информација овог извора је:

$$S = -\frac{1}{6}\log_2 \frac{1}{6} - \frac{5}{6}\log_2 \frac{5}{6} = 0{,}65$$

у битима. Низ од $n = 100$ бацања коцке се може реализовати на 2^{100} различитих начина. Међу таквима је и рецимо низ од 100 самих шестица који се скоро сигурно неће десити.

Међутим, највероватније је да ће реализовани низ имати око 100/6 шестица и око 500/6 бројева који нису шестице. Ових комбинација има укупно 2^{65}. То је последица закона великих бројева за Бернулијеву расподелу. ❑

82.2. Извор без меморије

Имамо најопштији низ независних случајних промењљивих $x_1, x_2, x_3, \dots$ од којих сваки x_i узима неку од вредности из скупа $\mathcal{A} = \{e_1, e_2, \dots, e_l\}$. Видели смо да је информација таквих речи дужине n

$$S_n = -\sum_{j=1}^{n} p_j \log_b p_j, \quad n = 1, 2, 3, \dots,$$

где је $p_j = \Pr\big(x_i = e_j\big)$, $i = 1, 2, 3, \dots$, $j = 1, 2, \dots, a$. Отуда је информација извора $S = S_n = S(x_i)$, тј. једнака је информацији једног елемента низа. Другим речима, S је средња вредност информације слова у низу.

До истог резултата долазимо полазећи од Хартлијеве дефиниције информације за једно слово

$$I = -\log_b \Pr(x_i),$$

па је очекивање:

$$EI = -\sum_{x_i \in \mathcal{A}} \Pr(x_i) \log_b \Pr(x_i) = -\sum_{j=1}^{n} p_j \log_b p_j = S.$$

То је средња информација једног слова азбуке. Средња информација емитоване поруке од n слова (међу којима може бити и једнаких) је:

$$I(x_1, \dots, x_n) = -\frac{1}{n}\log_b \Pr(x_1, \dots, x_n) = -\frac{1}{n}\sum_{i=1}^{n} \log_b \Pr(x_i),$$

јер извор нема меморију. Према томе је:

$$EI_n = \frac{1}{n}\sum_{i=1}^{n} EI = \frac{1}{n} \cdot n \cdot S = S.$$

Дакле, средња информација по слову азбуке једнака је средњој информацији слова поруке (независно од n) и једнака је информацији извора. Отуда је дисперзија константа извора, тачније

$$\sigma^2(I_n) = \sigma^2 I = \sigma^2 < \infty,$$

па $\frac{1}{n^2} \cdot \sigma^2 \to 0$, када $n \to \infty$, те извор без меморије испуњава услове Марковљевог закона великих бројева, тј.

$$\lim_{n \to \infty} \Pr\left(|I_n(\vec{x}) - S| \geq \varepsilon\right) = 0$$

за свако $\varepsilon > 0$. Према томе

$$b^{-n(S+\varepsilon)} \leq \Pr(\vec{x}) \leq b^{-n(S-\varepsilon)},$$

када $n \to \infty$. То значи да сваки низ $\vec{x}$ можемо раставити на два дела (комутација скуповног пресека) да је $\vec{x} = (\vec{x}_0, \vec{x}_1)$, па је:

$\Pr(\vec{x}) = \Pr(\vec{x}_0) \Pr(\vec{x}_1)$ – због независности,

$\Pr(\vec{x}_0) = 0, \quad \Pr(\vec{x}_1) = b^{-nS}.$

Тиме је језик растављен на два подскупа A_0 и A_1. Укупан број речи у A_1, опет због независности и због $\sum_{\vec{x} \in A_1} \Pr(\vec{x}) = 1$ је $\mathrm{No}(A_1, n) = b^{nS}$, где је $b > 1$ база логаритма у изразу за информацију. Кад су сва слова једнако вероватна, тада је информација највећа $S = \log_b l$, па $\mathrm{No}(A_1, n) = l^n$. То је укупни број комбинација дужине n од l слова и у том случају A_0 је празан скуп.

82.3. Марковљев ланац

Видели смо да је информација стационарног Марковљевог извора

$$S = -\sum_{i=1}^{l}\sum_{j=1}^{l} p_i \cdot p_{ij} \cdot \log p_{ij}.$$

Она се своди на претходни случај када је $p_{ij} = \Pr(e_j | e_i) = \Pr(e_j) = p_j$, тј. када су $x_1, x_2, x_3, \dots$ међу собно независни. Полазећи од условних вероватноћа имамо:

$$\Pr(x_1, \dots, x_n) = \Pr(x_1) + \Pr(x_2|x_1) + \cdots + \Pr(x_n|x_{n-1}),$$

$$I(\vec{x}) = -\frac{1}{n} \cdot \log \Pr(\vec{x}) = -\frac{1}{n} \cdot \log \Pr(x_1) - \frac{1}{n} \cdot \sum_{i=2}^{n} \log \Pr(x_i|x_{i-1}).$$

Отуда:

$$EI_n = -\frac{1}{n} \cdot E \log \Pr(x_1) - \frac{1}{n} \cdot \sum_{i=2}^{n} E \log \Pr(x_i|x_{i-1}) =$$

$$= -\frac{1}{n} \cdot S(x_1) - \frac{n-1}{n} \cdot \sum_{i,j=1}^{n} p_i \cdot p_j \cdot \log p_{ij} = S.$$

Дакле, средња информација поруке је S. Да бисмо избегли доказивање да $I_n \to S$, приметимо да низ случајних променљивих $\Pr(x_1)$, $\Pr(x_2|x_1)$, …, $\Pr(x_n|x_{n-1})$ узима вредности из коначног скупа од $l(l+1)$ елемената. Због стационарности ланца x-ова за $n \to \infty$ је низ ових вероватноћа извор без меморије. Према томе је и низ информација $I(x_1)$, $I(x_2|x_1)$, … $I(x_n|x_{n-1})$ извор без меморије. Отуда закон великих бројева

$$\lim_{n\to\infty} \Pr(|I_n - S| > \varepsilon) = 0$$

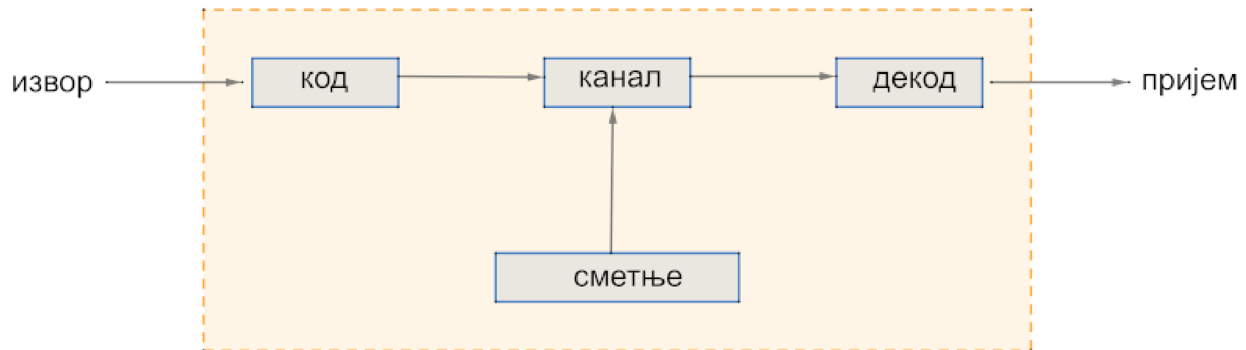
за свако $\varepsilon > 0$, где је

$$I_n = \frac{1}{n}[I(x_1) + I(x_2|x_1) + \cdots + I(x_n|x_{n-1})].$$

Даље растављање на скупове A_0 и A_1 је исто као у претходном случају.

83. Крипто код

Грубо речено, кодирање је намерна трансформација извора у облик из којег је могуће опет добити тачан полазни податак. Према томе, путујући кроз канал преноса кôд може стећи сметње које смо раније описивали, тако да грешке пријема неће бити грешке (де)кодирања. Такав пренос података је шематски приказан на следећој слици и тема је наставка текста.



Дакле, занемарујемо грешке оператера који раде на кодирању и декодирању које би било могуће учинити без грешака, ако би (кодирана) порука стизала без сметњи, и бавимо се само пригушеном информацијом насталом таквом лажирањем, привидним дезинформисањем, односно лагањем са намером обмане неовлаштеног читаоца, или његовог ометања.

Пример 1. Децимале броја $\pi = 3{,}14159265359 \dots$ када их први пут читамо изгледају као случајни бројеви. Оне ће проћи све тестове случајности математичке теорије вероватноће, а то сазнање је један од темеља моје (хипо)тезе о објективној случајности.

Подсећам, насумичност се заснива на немогућности поседовања свог знања одједном, било кога јединог субјекта, због чега ми комуницирамо. Због начелног минимализма (тежње ка мањој), информација послата од пошиљаоца прима се код примаоца због тог мањка. Она је на тај начин и слојевита, па можемо учити, стицати знање, тј. информације. Са друге стране, нека информација увек је од неких скривена, па су могуће обмане, рецимо ловца који свој плен лови клопком. Због истог, или боље речено доследно поставкама ове теорије информације, сама васиона нека је неизвесност и према својој унутрашњости и према спољашњости. Ово друго подразумева да ма како велику васиону успели сагледати, замислити, то неће бити све што постоји. ❑

Као и број π, децимале сваког ирационалног броја у првом читању изгледаће нам као „доказиво“ случајан низ бројева. Међутим, у сваком следећем читању истог броја приметићемо „превару“, тј. могућност да случајни бројеви за једног субјекта (нас у првом читању) нису случајни за неког другог субјекта (истих нас у другом читању). Овај смисао „објективне случајности“ експлоатисаћемо даље у тумачењу кодирања.

Претпостављамо и даље да имамо азбуку $\mathcal{A} = \{e_1, e_2, \dots, e_l\}$ од $l = 2, 3, \dots$ слова и извор података $x_0, x_1, x_2, \dots$ такав да је свако $x_i \in \mathcal{A}$. Међутим дефинишемо и помоћну азбуку $\pi = \{\pi_1, \pi_2, \dots, \pi_r\}$ од $r = 2, 3, \dots$ слова која опет, може произвести низ $y_0, y_1, y_2, \dots$ такав да је свако $y_j \in \pi$. Нај општије пресликавање низа x_i у низ y_j задајемо функцијом

$$f : \{x_{i_1}, x_{i_2}, \dots, x_{i_m}\} \to \{y_{j_1}, y_{j_2}, \dots, y_{j_n}\}$$

коју зовемо „кôд“ ако постоји поступак да се из низа y_j поново добије низ x_i. Обратни поступак се назива „декод“.

Пример 2. Најједноставнији случај кодирања је $m = n = 1$, $l = r$. Тада је функција кода задата са r једнакости $f(e_i) = \pi_j$, редом за $i, j = 1, 2, \dots, l$, или прегледније пресортирано

$$f = \begin{pmatrix} a & b & c & \dots \\ u & v & w & \dots \end{pmatrix}.$$

Горње слово азбуке $\mathcal{A}$ кодира се у доње слово азбуке π, а функција f је бијекција. Ако је кодиран низ $vuwwwu$ ... његов декод, оригинални низ био је $baccca$ ... . ❑

Наведени пример може бити користан једино дидактички. Са становишта брзине или сметњи тај тешко да може побољшати пренос података, а са становишта тајности података он је наиван. Али треба знати да се и најбољи крипто кодови (грч. ˈkrɪptəʊ – скривам, покривам) проваљују и да ће сваки важан податак бити дешифрован, само ако провалник има довољно времена. Ради тога је основни циљ криптографије (тајног писања) да време декодирања неовлаштеној особи учини што дужим, а да при томе овлаштеној особи декодирање не огорча живот.



Мало проширена метода овог примера је Цезарова шифра. То и јесте један од најпростијих и нај распрострањенијих начина кодирања, а врши се замењивањем сваког слова отвореног текста са одговарајућим словом друге азбуке. Тим дословним преписивањем датих слова неким другима, фреквенција првих у изворном тексту постаје неизмењена фреквенција у мењаном тексту. Како пад информације текста у односу на равномерну насумичност слова остаје сачуван, прикривање се лакше разоткрива. Слова фреквенције кодираног текста заменимо словима исте или најближе фреквенције изворне азбуке и направили смо велики корак у откривању поруке.

83.1. Вижнерова шифра

Много сложенија и боља метода шифровања од Цезарове је Вижнерова. Она се добија серијом замена Цезаровог шифрирања заснованих на словима кључа. У делу ове скрипте [61.3.] обрадио сам фреквенцију слова тајног текста са налазима који би овде могли бити корисни. Вижнеровим кодирањем подиже се информација, дакле смањује разлика информације датог кода и равномерно случајно нанизаних слова, што неовлаштеног читаоца додатно збуњује. Могућност дешифровања тада наликује могућности погађања децимала броја π, када смо број упознали. Псеудо случајност познаваоц преокреће у извесност.

У поменутом поднаслову [61.3. Вижнерова шифра] кодирање је помоћу “Vigenère standard code”, програма којег сам радио директно на интернет страници у php-у. Међутим, следећи два модула, или функције, су програми који раде то исто у језику “Python 3.9.7”. Рецимо текст “Konj pozoba sve

iz dzaka.", са кључем "ponedeljak", први од наведених модула "vgciph()" кодира (Encryption, encode) у " ZCAN SSKXBK HJR MC HKJKK.". Проверите!

```
def vgciph():
      tabula = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"
      print("Vigenère Cipher, coding")
      print(tabula)
      poruka1 = input("Plaintext: ")
      kljuc1 = input("Key: ")
      kljuc = kljuc1.upper()
      poruka = poruka1.upper()
      n = len(poruka)
      kljuc2 = ''

      while len(kljuc2) < n:
            kljuc2 = kljuc2 + kljuc
      tabula = tabula + tabula
      poruka2 = ''
      i = 0

      for x in poruka:
            x2 = x
            if tabula.find(x) > -1:
                  x1 = kljuc2[i]
                  i1 = tabula.find(x1)
                  i2 = tabula.find(x)
                  if (i1 != -1) and (i2 != -1):
                        x2 = tabula[i1+i2]
                  i = i + 1
            poruka2 = poruka2 + x2

      print(" Plaintext: ", poruka1)
      print("       Key: ", kljuc1)
      print("Ciphertext: ", poruka2)
```

Други од наведених модула "ciphvg()" декодира (Decryption, decode), враћа претходни код текста "zcan sskxbk hjr mc hkjkk.", у почетни "KONJ POZOBA SVE IZ DZAKA.", са истим кључем "ponedeljak". Пожељно је да тачност ових програма проверите на другим местима[22].

```
def ciphvg():
      tabula = "ABCDEFGHIJKLMNOPQRSTUVWXYZ"
      print("Vigenère Cipher, decoding")
      print(tabula)
      poruka1 = input("Ciphertext: ")
      kljuc1 = input("Key: ")
      kljuc = kljuc1.upper()
      poruka = poruka1.upper()
```

[22] Vigenere Tool, https://www.boxentriq.com/code-breaking/vigenere-cipher

```
n = len(poruka)
kljuc2 = ''

while len(kljuc2) < n:
      kljuc2 = kljuc2 + kljuc
tabula = tabula + tabula
poruka2 = ''
i = 0

for x in poruka:
      x2 = x
      if tabula.find(x) > -1:
            x1 = kljuc2[i]
            i1 = tabula.find(x1)
            i2 = tabula.find(x)
            if (i1 != -1) and (i2 != -1):
                  x2 = tabula[i2-i1]
            i = i + 1
      poruka2 = poruka2 + x2

print("Ciphertext: ", poruka1)
print("       Key: ", kljuc1)
print(" Plaintext: ", poruka2)
```

83.2. Скривање бројевима

Следећи пример је још једне занимљиве и познате методе шифровања, која слова замењује бројевима. Табела је (први) кључ који за своје потребе можете по потреби мењати.

	5	2	9	4	1	6	8	3	0	7
					I	L	O	R	U	X
2	A	C	E	G	J	M	P	S	V	Y
9	B	D	F	H	K	N	Q	T	W	Z
5	0	1	2	3	4	5	6	7	8	9

У овом случају, реч “KRIPTO” постаје број “913128938” који има јединствен декод. Добијени низ бројева даље сабирамо по модулу са случајним низом бројева (други кључ). Једноставнији од случајних низова су бинарни, нпр. 010011110…, па добијамо 923139048. Познајући случајни низ, операцију и полазну табелу можемо декодирати поруку.

Пример 3. Ево како би (Python 3.9.7) изгледао програм за ово кодирање (сакривање).

```
def sakrij():
    # azbuka
    slova = ['A','B','C','D','E','F','G','H','I','J','K','L','M',
             'N','O','P','Q','R','S','T','U','V','W','X','Y','Z']
    kljuc1 = ['25','95','22','92','29','99','24','94','1','21','91',
'6','26',96','8','28','98','3','23','93','0','20','90','7','27','97']
```

```
    kljuc2 = "010011110"
    print("Coding by numbers")
    poruka1 = input("Plaintext: ")
    poruka = poruka1.upper()
    print("Obraditi:", poruka)

    # prvi ključ
    lista = []
    for i,x in enumerate(poruka):
        y = x # y = ''
        if x in slova:
            j = slova.index(x)
            y = kljuc1[j]
        lista.append(y)
        continue
    poruka2 = ''.join(lista)
    print("1. kljuc:", poruka2)

    # drugi ključ
    n = len(poruka2)
    kljuc = kljuc2
    m = len(kljuc)
    while m < n:
            kljuc = kljuc + kljuc2
            m = len(kljuc)
    print(" + cifre:", kljuc)
    poruka = ''
    for i,x in enumerate(poruka2):
        if can_int(x):
            y1 = int(x)
            y2 = int(kljuc[i])
            y = y1 + y2
            z = str(y)
            z = z[-1]
        else: z = ' '
        poruka = poruka + z
        continue
    print("2. ključ:", poruka)
    return
```

За тестирање да ли текст (string) представља цели број (int) користим следећу функцију.

```
def can_int(string):
    try:
        int(string)
        return True
    except ValueError:
        return False
```

Ево једног примера покретања овог програма и екстремног текста, ради тестирања.

```
modul.sakrij()
Coding by numbers
Plaintext: Ćevapčići su u Šušnjarima!
Obraditi: ĆEVAPČIĆI SU U ŠUŠNJARIMA!
1. kljuc: Ć29202528Č1Ć1 230 0 Š0Š962125312625!
 + cifre: 010011110010011110010011110010011110
2. ključ:  39213638 2 1 341 0   0 073126313736
```

Слова унешене поруке која не припадају дефинисаној „азбуци“ преписују се таква каква су, због наредбе `y = x`, али ако ту наребу замените са `y = ''` тог преноса неће бити. ❑

Пример 4. Овако би изгледао програм за претходно декодирање (откривање).

```
def otkrij():
    # azbuka
    slova = ['A','B','C','D','E','F','G','H','I','J','K','L','M',

             'N','O','P','Q','R','S','T','U','V','W','X','Y','Z']

    kljuc1 = ['25','95','22','92','29','99','24','94','1','21','91',

'6','26','96','8','28','98','3','23','93','0','20','90','7','27','97']

    kljuc2 = "010011110"
    print("Coding by numbers")
    poruka2 = input("Coded text: ")
    print("Uneto je:", poruka2)

    # drugi ključ
    n = len(poruka2)
    kljuc = kljuc2
    m = len(kljuc)
    while m < n:
            kljuc = kljuc + kljuc2
            m = len(kljuc)
    print(" oduzeti:", kljuc)
    poruka = ''
    for i,x in enumerate(poruka2):
        if can_int(x):
            y1 = int(x) + 10
            y2 = int(kljuc[i])
            y = y1 - y2
            z = str(y)
            z = z[-1]
        else: z = ' '
        poruka = poruka + z
        continue
    print("Rezultat:", poruka)
```

```
    # prvi ključ
    lista = []
    z = ''
    for x in poruka:
        if can_int(x):
            z = z + x
        else:
            lista.append('.')
        if z in kljuc1:
            j = kljuc1.index(z)
            y = slova[j]
            lista.append(y)
            z = ''
        continue

    poruka1 = ''.join(lista)
    print("  Poruka:", poruka1)
    return
```

Тестирамо дешифровање (откривање) горњег кода:

```
modul.otkrij()
Coding by numbers
Coded text:  39213638 2 1 341 0  0 073126313736
Uneto je:  39213638 2 1 341 0  0 073126313736
 oduzeti: 01001111001001111001001111001001111 0
Rezultat:  29202528 1 1 230 0  0 962125312625
  Poruka: .EVAP.I.I.SU.U..U.NJARIMA.
```

Ознаку тачке (`'.'`) за код ван азбуке можете произвољно мењати. ❑

Пажљивим бирањем начина кодирања према садржају улазног текста може се веома ујединачити расподела цифара излазне шифре. То када фреквентнија слова шифрујемо са дужим бројевима, а мање фреквентна њиховим појединим цифрама. Тако повећан „шум“ неовлаштени би пријемник могао видети као „безначајан сигнал“ и одустати од дешифровања. Са друге стране, преласком са једне азбуке (са $l = 26$ слова) на другу (цифара са $r = 10$ слова) мења се распон опција и дужина кода (расте $l/r = 2{,}6$ пута). Колико смањивањем кодне азбуке добијемо на тајности или тачности кода и преноса, толико изгубимо на повећању дужине текста.

Пример 5. Размотримо опет исти мој текст „Феминизација“ из скрипте „Приче о информацији“, на енглеском, кориштен за израчунавање фреквенција слова [66.1. Енглески језик]. Тај текст:

```
Local and humorous in a private correspondence with colleagues I used
the term feminization for physical processes that give up of the
outside world. Look your own business, and don't worry for others...
```

кодиран истим програмом 3. примера постаје текст:

```
6922367 250693 05027839134 106 36 383220360429
22944392428907022096 3330 9119405 2296630362412934 2 0332003 930420
```

```
0429426 003927197208250319O7 9993 39052733133366 3839333924233034
94943604 2512130 029 800 039529 91032419330 9093603  68991 38903 89107
95123207392423  369602 03906 03 0194328 9094 8949430433...
```

чију фреквенцију слова (цифара) затим налазимо. Текст има 12518 цифара са фреквенцијама:

1.	3	2542	0,203
2.	0	2132	0,170
3.	9	2012	0,161
4.	2	1818	0,145
5.	4	1124	0,090
6.	6	784	0,063
7.	7	622	0,050
8.	1	552	0,044
9.	5	538	0,043
10.	8	394	0,031

Средња информација ове S'_y и максималне такве S'_0 расподеле је:

$$S'_y = -0{,}203 \cdot \log_2 0{,}203 - \cdots - 0{,}031 \cdot \log_2 0{,}031 = 2{,}961587 \ldots$$

$$S'_0 = -0{,}1 \cdot \log_2 0{,}1 - \cdots - 0{,}1 \cdot \log_2 0{,}1 = \log_2 10 = 3{,}321928 \ldots$$

Релативно велика разлика $S_0 - S_y > 0{,}360$ значи да скривени текст није значајно камуфлиран и са униформношћу расподеле слова. У тамошњем, изворном енглеском тексту [66.1], где је максимум информације $S_0 = \log_2 26 = 4{,}700440$ и овом, на шест децимала, коефицијенти шума су:

$$\eta = \frac{S_y}{S_0} = \frac{4{,}134927}{4{,}700440} = 0{,}879689 \quad \eta' = \frac{S'_y}{S'_0} = \frac{2{,}961587}{3{,}321928} = 0{,}891527$$

Дакле, $\eta < \eta'$ што значи да овај кôд ипак у некој мери чини камуфлажу голог текста. ❑

Када немамо бољи начин за добијање случајних низова (они генерисани рачунаром непоуздани су), користимо рекурентну релацију

$$y_{n+1} = (a \cdot y_n + b) \bmod m,$$

редом за $n = 0, 1, 2, \ldots$, где су a, b, m природни бројеви. Дефинишемо ли $a = 2$, $b = 3$ и $m = 5$, ако је $y_0 = 0$, добићемо $y_1 = 3$, $y_2 = 4$, $y_3 = 1$, $y_4 = 0$, $y_5 = 3$, … низ је даље периодичан. Рецимо, кад $y_0 = 2$ добијамо свако $y_n = 2$. Уопште, за свако су a, b, m и y_0, добијени низови су периодични са периодом не дужом од m.

Наведени пример схваћен као концепт са мало маште може се скоро неограничено проширивати. Следећи пример демонстрира једноставну употребу једне „заборављене“ азбуке која је згодна за скривање слова горње табеле, након чега можемо наставити кодирање (други кључ). Она чешћим словима даје краће кодове, што је „нелогично“ са становишта сакривања поруке, али је практично са становишта повећања капацитета преноса канала.

Пример 6. Заменимо „тачку“ са 0 и „црту“ са 1 Морзеове азбуке и добијамо:

A 01, B 1000, C 1010, D 100, E 0, F 0010, G 110, H 0000, I 00, J 0111, K 101, L 0100, M 11, N 10, O 111, P 0110, Q 1101, R 010, S 000, T 1, U 001, V 0001, W 011, X 1001, Y 1011, Z 1100

Ове замене користимо уместо горње табеле (први кључ), а затим наставимо као у том примеру. Резултат је код који познаваоц може декодирати. Када га обрадимо програмима 5. примера, исти енглески текст постаје низ:

```
02002222011011211 0210211 001001222112011222002001 1110 01
1220020011111021O 212012102112001001221112102111102011 1111011111
11102221101011111200121100 00 1120010111 200100 2111021
112101200211111100221111210 1120121021 ...
```

```
са укупно 18140 знакова (цифара 0, 1, или 2)фреквенција:
```

```
1. 1 9426  0,519
2. 0 4778  0,263
3. 2 3936  0,217
```

Средња информација ове S_y и максималне овакве S_0 расподеле је:

$$S"_y = -0{,}519 \cdot \log_2 0{,}519 - 0{,}263 \cdot \log_2 0{,}263 - 0{,}217 \cdot \log_2 0{,}217 \approx 1{,}476159$$

$$S"_0 = -\frac{1}{3} \cdot \log_2 \frac{1}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3} = \log_2 3 \approx 1{,}584962$$

Релативна разлика $S"_0 - S"_y = 0.109$ око 3,3 пута мања је од претходне, а отприлике толико пута је овај код дужи од претходног колики је и однос (10 : 3) броја цифара њихових заменских азбука. Међутим, однос дужина текстова два примера 18140 : 12518 = 1,449... показује да се Морзеовом заменом добија текст краћи него што би требао бити, што указује на већи капацитет таквог канала.

Захваљујући подређивању Морзеовог кода фреквенцији слова енглеског алфабета, он пребацује више интересантног текста по броју сигнала. Мања му је специфична информација ($S"_y < S'_y$), али са друге стране, коефицијенти шума оригиналног, претходног и овог кода:

$$\eta = 0{,}879689 \quad \eta' = 0{,}891527 \quad \eta" = \frac{1{,}476159}{1{,}584962} = 0{,}931353$$

у растућем су поретку η < η' < η". Кодирање и смањивање азбуке маскирају текст. ❑

Саме бројеве (други кључ) скривамо често помоћу прим-бројева (2, 3, 5, 7, 9, 11, 13, 17, 19, 23, ...). На пример, број 4596 израчунамо у број $2^4 \cdot 3^5 \cdot 5^9 \cdot 7^6 = 893397093750000$ који се може јединствено реконструисати растављањем на факторе у низ простих бројева. Замислите како овај посао може бити тежак неовлаштеном лицу ако низ простих бројева почиње негде у наставку, са веома великим бројевима. Растављање на факторе крајње великих бројева није лак посао ни најбољим рачунарима.

Пример 7. Следећи кораци чине један од ефикаснијих алгоритама за налажење простих бројева датог броја n.

1) Док год је n дељиво са 2, пишите 2 и делите n са 2.

2) После корака 1, n мора бити непарно. Тада започните петљу са индексом од $i = 3$ па до квадратног корена од n. Док год i дели n допишите i и поделите n са i. Након што n није више дељиво са иначе растућим индексом i, повећајте i за још 2 и наставите.
3) Ако је n прост број и већи је од 2, онда n неће постати 1 у претходна два корака. Дакле, одштампајте n ако је већи од 2.

Сваки сложени број има најмање један прост фактор мањи или једнак квадратном корену самог себе (тврдња корака 2). Наиме, нека је $a \cdot b = n$. Ако су оба фактора већа од $\sqrt{n}$, онда је $a \cdot b > n$, што је контрадикција са полазном претпоставком. Ево једног од програма који то ради.

```
def factor(n = 1):
    if not isinstance(n, int):
        print("Prirodni broj?")
        return
    if n < 1:
        print("Unesite pozitivan ceo broj!")
        return
    broj = n
    lista = []
    while broj % 2 == 0:
        lista.append(2)
        broj = broj//2
    import math
    m = int(math.sqrt(broj))
    for x in range(3, m, 2):
        while broj % x == 0:
            lista.append(x)
            broj = broj // x
    lista.append(broj)
    return lista
```

Модул тестирам различитим уносима.

```
modul.factor(69643379)
[13, 13, 19, 23, 23, 41, 1]
```

Према томе $69643379 = 13^2 \cdot 19 \cdot 23^2 \cdot 41 \cdot 1$, проверите! ❑

83.3. Метода остатка

Кинеска теорема о остатку каже да ако неко познаје остатке еуклидског дељења целог броја n са неколико целих бројева, онда се може јединствено одредити остатак дељења n производом ових целих бројева, под условом да су делиоци узајамно прости (било која два немају заједничких фактора осим јединице). Применићемо то у кодирању (другог кључа).

Означимо са $x_1 \cdots x_n$ низ који треба кодирати, са $y_1 \cdots y_n$ низ случајних бројева, а са $p_1 \cdots p_n$ низ узајамно простих бројева. Ознака $y||x$ је за операцију уланчавања (дописивања, надовезивања). Уланчавање обављамо у бинарном систему, тако да је нпр.

$$3\|1 = (011)_2\|(01)_2 = (01101)_2 = 13,$$

а нека је $p_1 = 14$ први већи број од добијеног 13 (узајамно прост са осталим p_1). Означавамо даље производе:

$$q = p_1 \cdots p_n, \quad q_1 = \frac{q}{p_1}, \dots, q_n = \frac{q}{p_n},$$

док је z_i решење једначине $q_i \cdot z_i = 1 \pmod q$. Коначно, кôд i-тог поља је $k_i = q_i \cdot z_i$, а интегрални кôд низа од n слова је збир

$$C = \sum_{i=1}^{n} k_i \cdot (y_i||x_i) \pmod q.$$

Познајући кôд C оригинални текст добијамо лакше због $y_i||x_i = C \pmod{p_i}$.

Пример 8. Кодирамо реч "IS", односно "123" горње табеле (први кључ). Нека је низ случајних бројева $y = 341$. Тада је:

$$3||1 = 01101_2 = 13 < 14 = p_1$$

$$4||2 = 10010_2 = 18 < 19 = p_2$$

$$1||3 = 00111_2 = 7 < 9 = p_3$$

Отуда $q = 14 \cdot 19 \cdot 9 = 2394$, $q_1 = 19 \cdot 9 = 171$, $q_2 = 14 \cdot 9 = 126$, $q_3 = 14 \cdot 19 = 266$, па:

$$171 \cdot z_1 = 1 \pmod{14} \to z_1 = 5$$

$$126 \cdot z_2 = 1 \pmod{19} \to z_2 = 8$$

$$266 \cdot z_3 = 1 \pmod 9 \to z_3 = 2$$

Коначно $k_1 = 171 \cdot 5 = 855$, $k_2 = 126 \cdot 8 = 1008$, $k_3 = 266 \cdot 2 = 532$, па је:

$$C = \sum_{i=1}^{3} k_i \cdot (y_i||x_i) \pmod q =$$

$$= (855 \cdot 13 + 1008 \cdot 18 + 532 \cdot 7) \pmod{2394}$$

$$= 32983 \bmod 2394 = 1861,$$

тј. кôд је $C = 1861$. Декодирањем добијамо:

$$1861 \bmod 14 = 13 = 3||1 \to 1$$

$$1861 \bmod 19 = 18 = 4||2 \to 2$$

$$1861 \bmod 9 = 7 = 1||3 \to 3$$

па је полазни низ $x = 123$, којем одговара реч "IS" табеле. ❑

84. Потпуност

Азбука је „потпуна“ ако садржи сва слова која су нам у тексту потребна. Тако је ASCII (American Standard Code for Information Interchange) табела потпуна ако користите само знакове (црвено) представљене у десним колонама следеће слике. У левим колонама су њихови кодови.

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040		Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174			|
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177	&#127;	DEL

Source: www.LookupTables.com

У првој левој колони стоји децимални кôд слова десно, у другој колони је хексадецимални кôд, у трећој html. На пример (65 41 101 A A), ако на рачунару држите типку Alt и при томе утипкате број 65, након пуштања прве типке (Alt-65) на екрану ће се појавити слово A. Свако велико и мало слово енглеског језика је у тој табели, као и цифре или неки типографски знаци често потребни.

Међутим, у горњој табели нема многих слова других језика, или рецимо знака „пуног квадратића“ који је ознака за крај доказа теореме. За неке такве постоји Продужена ASCII табела, приказана на следећој слици са словима (црвено) и децималним кодовима (црно). На пример, куцајте Alt-147 и приказаће се слово ô, дугосилазно акцентовано у српској речи кôд.

За куцање српских слова, латиничних đ, č, ć, š, ž и скоро свих ћириличних, обе те две табеле неће бити довољне. Као што знамо, за такве су смишљане тзв. кодне стране (Code Page), од којих је на персоналним рачунарима основна PC437, а касније још пуно других за разне језике. Тако постоје Latin1 (ISO-8859-1) за латинична писма Западне Европе (Француска, Немачка, Шпанија, ...), затим Latin2 (ISO-8859-2) и Windows-1250 за латинична писма Источне Европе (српска латиница и друге), потом ISO-8859-5, KOI8-R и Windows-1251 за ћирилицу. Да би се смањили проблеми настали код мешовитих текстова, уведен је Јуникод (Unicode), рецимо UTF-8 варијанте.

128	Ç	144	É	160	á	176	░	192	└	208	╨	224	α	240	≡
129	ü	145	æ	161	í	177	▒	193	┴	209	╤	225	ß	241	±
130	é	146	Æ	162	ó	178	▓	194	┬	210	╥	226	Γ	242	≥
131	â	147	ô	163	ú	179	│	195	├	211	╙	227	π	243	≤
132	ä	148	ö	164	ñ	180	┤	196	─	212	╘	228	Σ	244	⌠
133	à	149	ò	165	Ñ	181	╡	197	┼	213	╒	229	σ	245	⌡
134	å	150	û	166	ª	182	╢	198	╞	214	╓	230	µ	246	÷
135	ç	151	ù	167	º	183	╖	199	╟	215	╫	231	τ	247	≈
136	ê	152	ÿ	168	¿	184	╕	200	╚	216	╪	232	Φ	248	°
137	ë	153	Ö	169	⌐	185	╣	201	╔	217	┘	233	Θ	249	∙
138	è	154	Ü	170	¬	186	║	202	╩	218	┌	234	Ω	250	·
139	ï	155	¢	171	½	187	╗	203	╦	219	█	235	δ	251	√
140	î	156	£	172	¼	188	╝	204	╠	220	▄	236	∞	252	ⁿ
141	ì	157	¥	173	¡	189	╜	205	═	221	▌	237	φ	253	²
142	Ä	158	₧	174	«	190	╛	206	╬	222	▐	238	ε	254	■
143	Å	159	ƒ	175	»	191	┐	207	╧	223	▀	239	∩	255	

Source: www.LookupTables.com

То је први аспект појма „потпуности“ који нас занима. Огољена употреба азбука у комуникацији. А затим и проблеми са њиховом бројем, опширношћу и међусобним искључивањем, са навођењем нас да текстове пишемо истим писмом. Друга особина појма „потпуности“ тако нас слабо занима да је једва примећујемо и тешко ју је укратко објаснити, али вреди покушати.

Други аспект теме „потпуности“ је у самој суштини скривања, кодирања. Претпоставимо да нама примарна азбука $\mathcal{A} = \{a_1, \dots, a_l\}$ буде некоме наша секундарна заменска $\mathcal{B} = \{b_1, \dots, b_r\}$, држећи се претходног о кодирању. Бијекција текстова писаних са те две гарантује њихову еквивалентност. Да разумемо о чему говорим приметимо да кућни љубимац пас и мачка виде мање и слабије боје. Уопште, знамо за различите животиње које различито виде зраке светлости, боје спектра електро магнетног зрачења, а корак одатле је разумевање релативности примарних и секундарних азбука комуникација.

Кодиран и маскиран текст једном „читаоцу“ (субјекту интеракција), бар теоријски, можемо писати примарном азбуком неком другом. То прво зато што „разумевање“ поруке једном субјекту (атому, телу, особи) не мора то бити и неком другом. На пример, неутрони не реагују на „поруке“ електро магнетних поља као наелектрисани електрони и протони, рећи ћемо хипотетички, јер не разумеју језик наелектрисања. То је сасвим природна (хипо)теза ове „теорије информације“.

У том смислу теорија кодирања неће бити грана теорије информације, већ ће бити обрнуто, њена надградња. Погледајмо то на покушају интерпретације претходних преноса расподела каналима. Наиме, када горе поменуто кодирање $\mathcal{K} : \mathcal{A} \to \mathcal{B}$ пишемо матрично $\vec{b} = \hat{K}\vec{a}$, приметићемо да то некада има а некада нема смисла, иако постоји процес кодирања (вештачки или природни).

Рецимо да су слова енглеског алфабета управо ASCII децимални бројеви од 65 до 90, а кодирање дато табелом коју смо горе [83.2. Скривање бројевима] назвали првим кључем. То матрично је

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_{26} \end{pmatrix} = \begin{pmatrix} k_{1,1} & \cdots & k_{1,26} \\ \vdots & \ddots & \vdots \\ k_{26,1} & \cdots & k_{26,26} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{26} \end{pmatrix},$$

где су бројеви $a_1 = 65$, $a_2 = 66$, ..., $a_{26} = 90$ кодови редом великих слова енглеског алфабета, а компоненте $b_1 = 25$, $b_2 = 95$, ..., $b_{26} = 97$ њихови крипто кодови. Познајући коефицијенте матрице $\hat{K} = [\![k_{i,j}]\!]$ познајемо начин кодирања. Овако дефинисан процес крипто заштите може бити непрактичан, али није немогућ. Међутим, већ у следећи примерима тог наслова ова метода би запела. Не би било довољно ни увођење нелинеарних трансформација кодова, за сада још увек не прихватљиво у квантној физици. Зато ову идеју остављам „на чекању“.

84.1 Константне дужине

Ако азбука садржи 256 карактера тада је дужина кода једног карактера $\log_2 256 = 8$ бита једнаких дужина кодова. Одлучимо ли се за краћу азбуку од 128 карактера дужина једног кода биће 7 бита. Уопште смањимо ли неопходних слова два пута смањићемо трошење меморије за по један бит по слову оригинала, односно убрзати пренос за један бит. При томе је основно питање која од слова табеле задржати.

Дефиниција 1. Нека је $\mathcal{J}^m$ скуп свих речи дужине m изворног језика и нека је A^m скуп свих речи дужине m које се могу описати азбуком $\mathcal{A} = \{a_1, \dots, a_n\}$. Тада се вероватноћа:

$$p_m = \Pr(x \in A^m) = \mathrm{P}(A^m), \quad x \in \mathcal{J}^m$$

зове ниво потпуности или потпуност (и поузданост) дате азбуке за низове константне дужине m. ▨

Дакле, када бирамо азбуку $\mathcal{A}$ слова потребних високо вероватним речима језика $\mathcal{J}^m$ добићемо потпуност блиску јединици, па ћемо ретко бити у ситуацији да немамо слова за запис потребне речи. У рачунарској инсталацији је дужина кодиране речи извора $m = 1$, тј. једно слово, док је дужина кодне замене $n = 7$, за табелу од 128 слова.

Означимо ли са $\mathcal{K}^n$ скуп свих кодираних речи дужине n тада вероватноћу $P(\mathcal{K}^n)$ зовемо потпуност датог кода. Како је број свих речи дужине m над азбуком $\mathcal{A}$ једнак l^m, а број свих речи дужине n над азбуком $\mathcal{B}$ једнак r^n, тада је услов

$$l^m = r^n$$

потребан и двољан да је потпуност кода једнака потпуности азбуке. Ако је $l^m < r^n$ онда је могуће кодирати више речи него што је потребно за азбуку $\mathcal{A}$, па је опет $P(A^m) = P(\mathcal{K}^n)$ јер вишак кодних замена нема смисла декодирати. Обрнуто, $l^m > r^n$ значи да $P(A^m) > P(\mathcal{K}^n)$, тј. код има мању потпуност.

Без губитка општости можемо претпоставити да је потпуност основне азбуке један. У том случају можемо рећи да је $l^m \le r^n$ потребан и довољан услов да буде $P(\mathcal{K}^n) = 1$. Отуда

$$\frac{m}{n} \le \frac{\log r}{\log l},$$

ако и само ако $P(\mathcal{K}^n) = 1$. Ту је n/m број кодних симбола потребан за емитовање једног слова оригиналног извора података. Реципрочна вредност m/n је брзина преноса кода по слову извора.

84.2. Ергодичност

Са друге стране, када је извор $x_1, x_2, \ldots$ ергодички, могуће је изабрати за функцију кода (f) бијекцију такву да је потпуност кода, за произвољно $\varepsilon > 0$ једнака:

$$p_m = P(A^m) = \Pr(|I_m(\vec{x}) - S| \leq \varepsilon),$$

где је $\vec{x}$ подниз дужине m извора са информацијом $S > 0$. Знамо да је $I_m(\vec{x})$ просечна информација једног слова подниза $\vec{x}$. Због асимптотске дељивости (енг. AEP) ергодичког извора, тада је

$\lim_{m\to\infty} p_m = 1$.l

S Из истог разлога је тада $A^m = \{\vec{x} : |I_m(\vec{x}) - S| \leq \varepsilon\}$ високо вероватан подскуп језика $\mathcal{J}^m$. Из дефиниције информације $I_m(\vec{y})$ добијамо:

$$-m \cdot (S + \varepsilon) \leq \log_2 P(\vec{x}) \leq -m \cdot (S - \varepsilon),$$

$$2^{-m(S+\varepsilon)} \leq P(\vec{x}) \leq 2^{-m(S-\varepsilon)}.$$

Ако скуп A^m има M елемената биће:

$$M \cdot 2^{-m(S+\varepsilon)} \leq \sum_{\vec{x}\in A^m} P(\vec{x}) \leq 1,$$

$$M \leq 2^{m(S+\varepsilon)} < r^n,$$

јер је $l^m < r^n$. То је довољан услов да постоји инјекција

$$f : \{x_{i_1}, \ldots, x_{i_m}\} \to \{y_{j_1}, \ldots, y_{j_n}\}.$$

Довољан услов за егзистенцију кода (f) потпуности p_m за дати ергодични извор података је тада

$$\frac{n}{m} > \frac{S + \varepsilon}{\log l}. \quad (*)$$

Дакле, повећавањем броја кодних симбола по једном слову оригинала, тј. за велико m је увек могуће конструисати кôд са поузданошћу близу један и са брзином преноса (m/n) близу $\log l / S$. Обратно, ако је

$$\frac{n}{m} \leq \frac{S - 2\varepsilon}{\log l},$$

за произвољно мало $\varepsilon > 0$, тада је за велико m потпуност кода произвољно блиска нули. Наиме, чак и ако је то високо вероватан подскуп речи, из:

$$n \log l \leq m \cdot (S - 2\varepsilon) \iff l^n \leq 2^{m(S-2\varepsilon)}$$

бидимо да број елемената скупа $\mathcal{K}^n$ није већи од $2^{m(S-2\varepsilon)}$ и да вероватноћа сваког појединачног елемента не прелази:

$$2^{m(S-2\varepsilon)} \cdot 2^{-m(S-\varepsilon)} = 2^{-m\varepsilon} \to 0,$$

када $m \to \infty$.

Закључујући, можемо рећи да је услов (*), уз $m \to \infty$ и ергодичност извора потребан и довољан за постојање поузданог кода, тј. таквог кода који са вероватноћом један може имати управо ону реч која нам је потребна. Међутим, закони великих бројева о којима се ради не важе за мале бројеве m. То је један од најважнијих разлога због којих се уводе кодови са променљивом дужином кодних замена.

84.3. Јединственост

Кодови са сталном дужином кодних замена, које смо управо расправљали, веома успоравају пренос података и непотпуни су. То превазилазимо помоћу брзих процесора и алтернативним азбукама, односно брзим меморијама и једноставнијим алгоритмима декодирања. Насупрот њима, кодови са променљивом дужином кодних замена немају проблема са брзином и потпуношћу, али имају дотле непознате проблеме са једнозначном препознатљивошћу, тј. са јединственошћу.

Када се не бавимо и маскирањем (кодирањем налик околини), скуп свих речи поредајмо по вероватноћи њиховог појављивања у језику, затим највероватнијим придружимо најкраће кодне замене. Тиме добијамо кôд са максималном потпуношћу и великом брзином. Са друге стране, такво придруживање је само у изузетним случајевима могуће декодирати, а наш је задатак да утврдимо те изузетне ситуације.

Пример 1. Дате су азбуке $\mathcal{A} = \{a_1, a_2, a_3\}$, $\mathcal{B} = \{0, 1\}$ са познатим вероватноћама $P(a_1) = 0{,}5$, $P(a_2) = 0{,}3$, $P(a_3) = 0{,}2$. Ако дефинишемо придруживање $f\,:\ a_1 \to 0, a_2 \to 1, a_3 \to 01$, видимо да f није прави кôд, јер рецимо након емисије $a_1a_2a_3$ добијамо низ 0101. Онај који поруку декодира овај низ може схватити као $a_1a_2a_1a_2$ или $a_3a_1a_2$ или a_3a_3. Преме томе, дефинисани „кôд“ нема јединствен декод. ❑

Пример 2. Имамо исте азбуке и исте вероватноће слова као у претходном примеру али их кодирамо другачије $f\,:\ a_1 \to 1, a_2 \to 10, a_3 \to 100$. Након емисије нпр. a_1a_2 декодер прима поруку 110 која се може једнозначно дешифровати али само под условом да је пренос завршен. Наиме, прими ли декодер четврту цифру 0 онда је изворни низ a_1a_3. ❑

Пример 3. Под истим условима дефинишемо код $f\,:\ a_1 \to 0, a_2 \to 10, a_3 \to 11$. Изворни низ $a_1a_2a_3$ постаје 01011, а овај се јединствено декодира без обзира на крај преноса. ❑

Означимо са $\mathcal{K} = f(\mathcal{J}_A)$ скуп свих речи азбуке $\mathcal{B}$ које можемо добити пресликавањем f извора азбуке $\mathcal{A} = \{a_1, a_2, \dots, a_l\}$. Специјално $k_i = f(a_i)$, $i = 1, 2, \dots, l$ замена слова азбуке $\mathcal{A}$ или другим речима елементарна реч азбуке $\mathcal{B} = \{b_1, b_2, \dots, b_r\}$ која је састављена из једног или више слова те кодне азбуке. Јасно је да сваки пар низова $x, y\ \in \mathcal{K}$ ту је и продужени низ $z = x\|y \in \mathcal{K}$. Такође, за свако $x, y, z \in \mathcal{K}$ је $x\|(y\|z) = (x\|y)\|z$, тј. вреди закон асоцијације. Са затвореношћу и асоцијативношћу скуп $\mathcal{K}$ у који је уведена операција уланчавања ‖ чини полугрупу.

Ако $x, y \in \mathcal{K}$, тада се x зове префикс а y суфикс речи $z = x\|y$. Даље је $\mathcal{K}_0 = \{f(a_1), \dots, f(a_l)\}$ скуп свих замена, а $\mathcal{K}_1 = \{y \in \mathcal{K} : x\|y \in \mathcal{K}_0, x \in \mathcal{K}_0\}$ скуп свих суфикса елементарних речи са префиксима такође елементарним речима. Формирајмо даље уније $\mathcal{K}_j = \mathcal{K}'_j \cup \mathcal{K}"_j$ редом за $j = 1, 2, 3, \dots$, где су $\mathcal{K}'_j = \{y \in \mathcal{K} : x\|y \in \mathcal{K}_{j-1}, x \in \mathcal{K}_0\}$ и $\mathcal{K}"_j = \{y \in \mathcal{K} : x\|y \in \mathcal{K}_0, x \in \mathcal{K}_{j-1}\}$.

Приметимо да је сваки од скупова $\mathcal{K}_1, \mathcal{K}_2, \dots$ састављен од суфикса којима је префикс или елементарна реч ($\mathcal{K}_0$) или елемент скупа из поменутог низа. Како је $\mathcal{K}_0$ коначан скуп, а његови елементи

имају коначну дужину, то је и укупан број суфикса коначан. Према томе, у низу $\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2, \dots$ има само коначно много различитих скупова, па скуп свих суфикса $C = \mathcal{K}_1 \cup \mathcal{K}_2 \cup \dots$ такође има коначно много елемената.

Пример 4. У првом примеру је $\mathcal{K}_0 = \{0, 1, 01\}$ и $\mathcal{K}_1 = \{1\}$, док су сви остали скупови празни. Тако је $C = \mathcal{K}_1$ па је $\mathcal{K}_1 \cap C = \{1\}$.

У другом примеру су $\mathcal{K}_0 = \{1, 10, 100\}$ и $\mathcal{K}_1 = \{0, 00\}$ једини непразни скупови, па је $C = \mathcal{K}_1$ и $\mathcal{K}_0 \cap C = \emptyset$.

У трећем примеру је $\mathcal{K}_0 = \{0, 10, 11\}$, а сви остали скупови $\mathcal{K}_j$ су празни. ❑

Када елементарне речи немају суфикса, тј. када је $\mathcal{K}_1 = \emptyset$, тада за декодирање очито није битан крај преноса података, па се такви кодови зову тренутни, или кодови са особином префикса.

Теорема 1. Инјекција f има јединствен декод само ако је $\mathcal{K}_1 \cap C = \emptyset$, тј. ако не постоји суфикс који је уједно елементарна реч.

Доказ: Претпоставимо супротно и узмимо најкраћи неједнозначан низ:

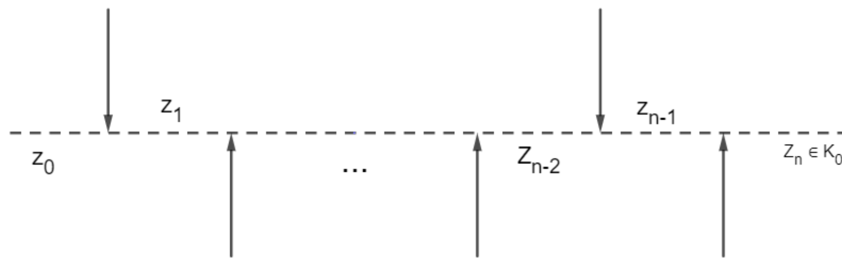
$$z = x_1 \| x_2 \| \dots x_\alpha = y_1 \| y_2 \| \dots y_\beta.$$

То значи да је z слика два различита низа слова азбуке $\mathcal{A}$, али је

$$f^{-1}(x_\alpha) \neq f^{-1}(y_\beta).$$

Тада је специјално x_α суфикс од y_β или обрнуто. Према томе, постоји суфикс који је замена што противречи претпоставци става.

Обрнуто, ако је $\mathcal{K}_0 \cap C \neq \emptyset$, онда постоји суфикс који је елементарна реч $k \in \mathcal{K}_j$. Тада се у највише коначно корака, унатрашке може конструисати низ $z \in \mathcal{K}$ који је могуће двозначно дешифровати:



Тиме је теорема доказана. ■

Непосредна последица овог става је јединственост декодирања сваког тренутног кода. На пример, Морзеова азбука није тренутни код, јер EA = (0)(01) = (00)(1) = IT и обе речи (EA, IT) кодирају се са иста три сигнала „тачка-тачка-црта“ (001).

85. Оптималност

Оптимална брзина кода је постигнута када је просечна дужина кодних замена минимална у скупу свих тренутних кодова. Просечну дужину кодне замене, за дати код дефинишемо као средњу вредност, односно матемаичко очекивање дужине елементарне речи

$$\bar{n} = En = \sum_{j=1}^{l} p_j n_j\,, \quad l = 2,3,\dots$$

где је $p_j = \Pr(a_j)$, n_j је број слова кода a_j. Азбука је $\mathcal{A} = \{a_1, a_2, \dots, a_l\}$ и претпостављамо да се пренос једног слова обавља у јединици времена, па је $\bar{n}$ очекивано време преноса кодне речи, а брзина преноса минимална када је просечна дужина елементарне речи минимална. Конкретно:

1. пример [84.3. Јединственост] даје $\bar{n} = 0{,}5 \cdot 1 + 0{,}3 \cdot 1 + 0{,}2 \cdot 2 = 1{,}2$;
2. пример даје $\bar{n} = 0{,}5 \cdot 1 + 0{,}3 \cdot 2 + 0{,}2 \cdot 3 = 1{,}7$;
3. пример, тренутни код $\bar{n} = 0{,}5 \cdot 1 + 0{,}3 \cdot 2 + 0{,}2 \cdot 2 = 1{,}5$.

Уопште, када се k-то слово азбуке $\mathcal{A} = \{a_1, a_2, \dots, a_l\}$ кодира са n_k слова $\mathcal{B} = \{b_1, b_2, \dots, b_r\}$ начина таквог кодирања је r^{n_k}, а вероватноћа избора једног од њих је r^{-n_k}. На пример, бинарно кодирање ($r = 2$) са три знака ($n_k = 3$) има осам начина ($2^3 = 8$). Тачно колико и троцифрених бинарних бројева (000, 001, 010, 011, 100, 101, 110, 111), са вероватноћом $2^{-3} = 0{,}125$ избора појединог.

Када би свако од кодираних слова ($k = 1, \dots, l$) имало наведену особину, а поменуте вероватноће испуњавале услов расподеле да је њихов збир један (или мањи), тада би, бар у теорији, постојале комбинације да „догађаји“ које представљају кодирана слова буду исходи расподеле. Видећемо да је то заиста одлика јединственог кодирања, а пре тога погледајмо још један пример.

Пример 1. Дата су азбука $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$ и њен код $\mathcal{B} = \{b_1, b_2, b_3, b_4\}$ табелом:

A	a_1	a_2	a_3	a_4
B	1	10	101	111
$\Pr(a_i)$	0,4	0,3	0,2	0,1

Имамо тренутни код очекиване дужине слова $\bar{n} = 0{,}4 \cdot 1 + 0{,}3 \cdot 2 + 0{,}2 \cdot 3 + 0{,}1 \cdot 3 = 1{,}9$. Како су вероватноће појаве слова у тексту опадајуће са порастом дужине кôда, овај кôд је оптималан. ❑

Примерима приметите разлике између максималне брзине кода, оптималности и јединствености. Посебно, кодна азбука, $\mathcal{B} = f(\mathcal{A})$, из последњег примера бинарна је ($r = 2$). Дужине кодова слова редом су $n_1 = 1, n_2 = 2, n_3 = 3$ и $n_4 = 3$, а поменути збир вероватноћа $2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1$. Ове посебне вероватноће, дакле, могу представљати неку расподелу, независне и потпуне исходе из неког скупа, због чега су биле могуће дате комбинације са којима је кôд јединствен. Ово интуитивно запажање о „посебним вероватноћама“ могуће је строжије исказати.

Теорема 1. (Крафтова неједнакост) Важи неједнакост

$$\sum_{k=1}^{l} r^{-n_k} \le 1$$

ако и само ако је код f јединствен.

Доказ: Означимо са $N = \max\{n_1, \dots, n_l\}$, а са η_j број кодних замена дужине $j = 1, 2, \dots, N$. Бирамо произвољан природан број k па је:

$$\left(\sum_{i=1}^{l} r^{-n_i}\right)^k = \left(\sum_{j=1}^{N} \eta_j r^{-j}\right)^k = \sum_{j=k}^{k-N} \beta_j r^{-j}$$

где је $\beta_j = \sum_{j_1+\dots+j_k=j} \eta_{j_1} \dots \eta_{j_k}$. Прва једнакост је тривијална, а другу добијамо степеновањем и сређивањем израза. Свака кодна замена дужине j_1 се комбинује са сваком кодном заменом дужине $j_2, j_3, \dots$ и добија j-члани низ кодне азбуке. Дакле, β_j је број различитих низова слова азбуке $\mathcal{A}$ који се кодирањем могу превести у j-члани низ слова кодне азбуке $\mathcal{B}$. Уопште, j-чланих низова азбуке $\mathcal{B}$ има највише r^j, па услов $\beta_j > r^j$ значи да декод није јединствен.

Ако је декод (f) јединствен биће $\beta_j \le r^j$, те:

$$\sum_{i=1}^{l} r^{-n_j} \le \left(\sum_{j=k}^{k-N} r^j r^{-j}\right)^{\frac{1}{k}} = (kN - k + 1)^{\frac{1}{k}} \to 1,$$

кад $k \to \infty$, што можемо провести јер је k произвољно. Обратно, ако не важи неједнакост твр-ђења, бар за једно j добијамо $\beta_j > r^j$, што значи да декод није јединствен. Тиме је теорема до-казана у потпуности. ■

Пример 1. У примерима из претходног наслова имамо:

1. $n_1 = 1, n_2 = 1, n_3 = 2$, па је $2^{-1} + 2^{-1} + 2^{-2} = 1{,}25 > 1$;
2. $n_1 = 1, n_2 = 2, n_3 = 3$, па је $2^{-1} + 2^{-2} + 2^{-3} = 0{,}875 < 1$;
3. $n_1 = 1, n_2 = 2, n_3 = 2$, па је $2^{-1} + 2^{-2} + 2^{-2} = 1 \le 1$.

У трећем случају кôд је био без суфикса у скупу елементарних кодних речи, тј. тренутни код који је специјални случај правог кода и за који такође вреди Крафтова неједнакост. ❑

Ако је испуњена Крафтова неједнакост (теорема 1) онда је увек могуће конструисати тренутни кôд са истим дужинама кодних речи $n_1, \dots, n_r$. Према томе, оптималне кодове можемо увек тражити у скупу тренутних. Тако у 2. примеру прошлог наслова [84.3.] уместо {1, 10, 100} имамо тренутни кôд {0, 10, 110}, или {1, 01, 001}.

Теоријски, могуће је једнозначно искодирати сву васиону, бесконачно много честица (фермиона), кодовима дужина природних бојева ($n_1 = 1, n_2 = 2, n_3 = 3, \dots$) самим бинарним цифрама 0 или 1, јер је Крафтов збир $2^{-1} + 2^{-2} + 2^{-3} + \dots = 1$. Штавише, на начин писања реалних бројева могуће је кодирање и континуума различитих појмова на једнозначне начине.

Теорема 2. (Шенон-Фано) Нека је дат дискретни стационарни извор података у којем слово $a_i \in \mathcal{A}$ има вероватноћу $p_i \ge 0, p_1 + p_2 + \dots + p_l = 1$.

1) Ако је $S_1 = -\sum_{i=1}^{l} p_i \log p_i$, тада постоји тренутни код за који вреди $\frac{S_1}{\log r} \le \bar{n} < \frac{S_1}{\log r} + 1$.

2) Ако је S информација извора, а $\bar{n}/k$ просечан број кодних симбола по једном симболу извора, тада је $\lim_{k\to\infty} \frac{\bar{n}_k}{k} = \frac{N}{\log r}$.

Доказ: Одредимо l целих бројева $n_1, n_2, \dots, n_l$ тако да је $r^{-n_i} \le p_i \le r^{-n_i+1}$, $i = 1, 2, \dots, l$. Тада је

$$\sum_{i=1}^{l} r^{-n_i} \le \sum_{i=1}^{l} p_i = 1,$$

па је могуће конструисати тренутни кôд са дужинама кодних замена $n_1, n_2, \dots, n_l$. Такође је:

$$\log p_i < (-n_i + 1) \log r,$$

$$\sum_{i=1}^{l} p_i \log p_i < -\left(\sum_{i=1}^{l} p_i n_i - 1\right) \log r = (\bar{n} - 1) \log r,$$

$$\bar{n} < \frac{S_1}{\log r} + 1,$$

а то је друга неједнакост из (1). Сменом $q_i = r^{-n_i}$ добијамо прву неједнакост, при чему једнакост вреди ако и само ако је $r^{-n_i} = p_i$ за свако $i = 1, 2, \dots, l$. Тиме је доказано (1).

Како је $S(x_1, \dots, x_k) = kS$ биће даље:

$$\frac{S(x_1 \dots, x_k)}{\log r} \le \bar{n}_k < \frac{S(x_1, \dots, x_k)}{\log r} + 1,$$

$$\frac{S(x_1 \dots, x_k)}{k} \cdot \frac{1}{\log r} \le \frac{\bar{n}_k}{k} < \frac{S(x_1 \dots, x_k)}{k} \cdot \frac{1}{\log r} + \frac{1}{k}.$$

Преласком на лимес, када $k \to \infty$, произилази (2). Тиме је теорема доказана у потпуности. ■

Другим речима, доказали смо да је $S/\log r$ оптимална дужина кодне замене за произвољан стационаран извор података информације S, односно да је то минимум у скупу тренутних кодова азбуке $\mathcal{B} = \{b_1, \dots, b_r\}$. У 1. примеру биће:

$$\frac{S}{\log r} = \frac{-0{,}4 \cdot \log_2 0{,}4 - \dots - 0{,}1 \cdot \log_2 0{,}1}{\log_2 2} = \frac{1{,}846439 \dots}{1} \approx 1{,}9 = \bar{n}.$$

Број $S/\log r$ исти је у било којој бази логаритма, због начина промене базе.

Прва неједнакост 2. теореме (1) је први део Шенонове (1948) теореме кодирања у којој он уместо овде информације (S) говори о ентропији (ознаке H), што се још увек ради у званичној теорији. Те забуне држао сам се и ја у својим текстовима крајем 20. века[23], све до пажљиве анализе ентропије у термодинамици. Много сам о томе писао у међувремену, па и популарно у недавном блогу.

[23] Растко Вуковић: *Математичка теорије информације и комуникације*, Друштво математичара Републике Српске, Бања Лука 1995.

Док ентропија супстанце расте, њена информација опада. Спонтани раст ентропије, којим топлота са тела веће температуре прелази на суседно тело мање температуре, прати смањење осцилација молекула које се хладе и мања укупна информација тела. Према томе, постоји спонтано опадање информације тела, праћено смиривањем хаотичног микро-кретања, а оно је у складу са начелним чешћим реализовањем вероватнијих исхода. Вероватнији догађаји мање су информативни.

***

Питање: У „теорији информација" све што је истинито стварно је, и некако „реално" се дешава?

Одговор: Да, прво хипотетички, а затим ми је то постајала крупна, озбиљна и забавна прича. На пример, васиона садржи огроман број честица и захтева неограничено растући број њихових положаја. Оваква „теорија информације" не дозвољава дословно понављање садашњости, па онда постоји и толики број једнозначних њихових „означавања" (кодова, шифровања, в. горње теореме и описе).

Ако првима „реалностима" треба пребројива бесконачност, онда другима треба непребројива неких. Међутим, ми већ имамо такав континуум реалних бројева, а намеће се питање - где су они? Наравно, појам „где" не мора одређивати само физичко место, па даље радимо и са апстрактним математичким ,,просторима''. Такви се даље надовезују и ова прича добија смисао ,,свеприсутног ткива информације у простору, времену и материји, и даље, чија суштина је неизвесност".

У таквој, васиона је такође објекат, па и неизвесност ка унутра и ка вани. А то је доследно Раселовом парадоксу (нема скупа свих скупова), или Геделовој теореми немогућности (не постоји теорија свих теорија), за разлику од свих других данас познатих „теорија реалности". Зато не одустајем од „теорије информације". А то је, приметимо, само један од начина како ова теорија долази до тих познатих немогућности.

Питање: Нарастање информације васионе памћењем и гомилањем прошлости нарушава ли закон одржања информације?

Одговор: Не, уколико се супстанца садашњости проређује, а њена укупна информација смањује за тачно ону количину до које допире сећање (Growing). Информација садашњости опада, ентропија расте и садашњост се хлади.

86. Хуфманов алгоритам

Завршићемо основна разматрања о коду са промењивом дужином кодних замена наводећи један познати алгоритам за конструкцију оптималног тренутног кода. Пре тога погледајмо доказе за два иначе интуитивно јасна става.

Теорема 1. У оптималном коду (f) слова веће вероватноће имају краће кодне замене. Прецизније

$$p_i > p_j \;\Rightarrow\; n_i < n_j$$

за све $i, j \in \{1, \dots, r\}$.

Доказ: Претпоставимо супротно, да постоје индекси $i, j \in \{1, \dots, r\}$ када $p_i > p_j \;\Rightarrow\; n_i < n_j$. Конструишемо кôд ($g$):

$$g(a_k) = \begin{cases} f(a_k), & k \notin \{i, j\} \\ f(a_j), & k = i \\ f(a_i), & k = j \end{cases}$$

Дакле, у кодовима f и g замењене су кодне замене a_i, a_j, а све остале су им остале исте. Даље, из:

$$0 < (p_i - p_j)(n_i - n_j) = p_i n_i - p_i n_j - p_j n_i + p_j n_j$$

следи $p_i n_j + p_j n_i < p_i n_i + p_j n_j$, тј. $\bar{n}(g) < \bar{n}(f)$, а то је контрадикција. Тиме је став доказан. ∎

Једноставно речено, спајајући коефицијенте „већи са мањим" (и мањи са већим) добијамо мањи резултат $S = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = \vec{x} \cdot \vec{y}$ скаларног множења два вектора $\vec{x} = (x_1, x_2, \dots, x_n)$ и $\vec{y} = (y_1, y_2, \dots, y_n)$. Обрнуто, спајајући „већи са већим" (мањи са мањим) добијамо већи резултат. То су познате особине „информације перцепције", важне у развоју ове теорије информације.

Сада приметимо да мања „информација перцепције" (њена форма у израчунавању „просечне" дужине кода), значи већу ефикасност у смислу брзине преноса података. Аналогна су запажања о већој ефикасности у смислу уређивања друштва, да она такође значе мање „непотребно расипање" снага, као и смањивање количине „непотребних опција" у смисленијем говору. Сва три, свака на свој начи, говори заправо о смањивању виталности, односно слобода, или информације.

Теорема 2. Постоји оптимални тренутни кôд (f) такав да кодне замене $f(a_{r-l}), \dots, f(a_l)$ имају исту дужину $n_{l-k} = \dots = n_l$ а разликују се само у последњем кодном слову. При томе је $k = r - 1$, ако је $l - 1$ дељиво са $r - 1$, а иначе је k остатак дељења $l - 1$ са $r - 1$.

Доказ: Ако је кôд оптималан, због 1. теореме је $n_{l-k} \le \dots < n_l$. Ако бар на једном месту вреди строга неједнакост, тада узимамо префикс дужине $n_{l-k} - 1$ кодне замене $f(a_{l-k})$ који уланчавамо редом кодним словима $b_1, \dots, b_{k+1}$ те добијамо нове кодне замене слова $a_{l-k}, \dots, a_l$, тј. нови код g. Тада је $n_l(g) < n_l(f)$, тј. $\bar{n}(g) < \bar{n}(f)$ што је у контрадикцији са претпоставком става, да је f оптималан кôд. Тиме је став доказан. ∎

Кôд из прошлог наслова [84. Пример 1], где се четири слова азбуке $\mathcal{A}$ пресликавају редом у 1, 10, 101, 111, са бројем слова $l = 4$ и $r = 2$, има исту дужину (3) последње две замене, те је $k = 1$. При томе, управо како каже теорема, је $k = r - 1 = 1$, јер је $l - 1 = 3$ дељиво са $r - 1$.

Са последње две теореме имамо све потребне кораке за конструкцију оптималног тренутног кода познатог као Хуфманов (Huffman, 1952) алгоритам. Потсетимо се, радимо кодирање прве азбуке $\mathcal{A} = \{a_1, a_2, \dots, a_l\}$ помоћу слова друге азбуке $\mathcal{B} = \{b_1, b_2, \dots, b_r\}$ функцијом $f(a_i) = b_{i_1} \dots b_{i_n}$, при чему је вероватноћа i-тог слова прве азбуке p_i. Хуфманови кораци су:

1. азбуку $\mathcal{A}$ сортирати тако да је $p_1 \ge p_2 \ge \cdots \ge p_l$;
2. проверити да ли постоји природан број k такав да је $l = k(r-1) + r$, због теореме 2.
3. ако не постоји такав број k онда азбуци $\mathcal{A}$ додати толико фиктивних слова, са вероватноћом нула, колико је потребно да број k постоји; специјално за $r = 2$ (бинарни кôд) увек постоји такво k да је $l = k + 2$, па азбуку $\mathcal{A}$ не треба проширивати;
4. формирати нову азбуку $\mathcal{A}_1$ са l_1 слова удруживањем последњих r слова претходне азбуке $\mathcal{A}$ у једно слово са вероватноћом која је збир вероватноћа здружених слова. Сваком од r здружених слова се придружи по једно од r слова кодне азбуке $\mathcal{B}$;
5. понављати претходни корак (4) све до $l_1 = r$.

Пример 1. Нека је $l = 5$, $r = 2$, те $\mathcal{B} = \{0, 1\}$. Кораци алгоритма су:

1) $p = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ 0{,}4 & 0{,}3 & 0{,}2 & 0{,}05 & 0{,}05 \end{pmatrix}$;
2) постоји $k = 3$ такво да је $5 = 3 \cdot (2-1) + 2$;
3) непотребно;
4) $p^{(1)} = \begin{pmatrix} a_1 & a_2 & a_3 & a_{45} \\ 0{,}4 & 0{,}3 & 0{,}2 & 0{,}1 \end{pmatrix}$, $a_4 \leftrightarrow 0$, $a_5 \leftrightarrow 1$;
   $p^{(2)} = \begin{pmatrix} a_1 & a_2 & a_{345} \\ 0{,}4 & 0{,}3 & 0{,}3 \end{pmatrix}$, $a_3 \leftrightarrow 0$, $a_{45} \leftrightarrow 1$;
   $p^{(3)} = \begin{pmatrix} a_{2345} & a_1 \\ 0{,}6 & 0{,}4 \end{pmatrix}$, $a_2 \leftrightarrow 0$, $a_{345} \leftrightarrow 1$.

Коначно $a_{12345} \leftrightarrow 0$, па имамо тренутни кôд

$$f = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \\ 0 & 10 & 110 & 1110 & 1111 \end{pmatrix},$$

са средњом дужином кодираног слова $\bar{n} = 2$. ❏

Хуфманов алгоритам за $r = 2$, једноставно речено, ће након прва три корака делити азбуку у две групе, првој дописати први а другој други кôд и понављати. Заправо, у обрнутом току.

Пример 2. Додајемо нуле и јединице доследно објашњењу и поново сортирамо:

$$\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 0{,}40 & 0{,}25 & 0{,}15 & 0{,}10 & 0{,}06 & 0{,}04 \end{pmatrix} \to \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_{56} \\ 0{,}40 & 0{,}25 & 0{,}15 & 0{,}10 & 0{,}10 \end{pmatrix} \to$$

$$\to \begin{pmatrix} a_1 & a_2 & a_{456} & a_3 \\ 0{,}40 & 0{,}25 & 0{,}20 & 0{,}15 \end{pmatrix} \to \begin{pmatrix} a_1 & a_{3456} & a_2 \\ 0{,}40 & 0{,}35 & 0{,}25 \end{pmatrix} \to \begin{pmatrix} a_{23456} & a_1 \\ 0{,}60 & 0{,}40 \end{pmatrix}$$

	a_5	a_6			
a_4	0	1			
0	10	11	a_3		
00	010	011	1	a_2	
000	0010	0011	01	1	a_1
0000	00010	00011	001	01	1

Слова азбуке сачувала су своје вероватноће. ❑

86.1. Рекапитулација

Кроз један пример, кодирањима исте поруке „BCCABBDDAECCBBAEDDCC“, поновимо неке главне моменте расправљане током последњих неколико наслова. Порука од 20 слова састоји се од само пет слова ASCII кодова наведених децимално и бинарно и приказаних у следећој табели[24].

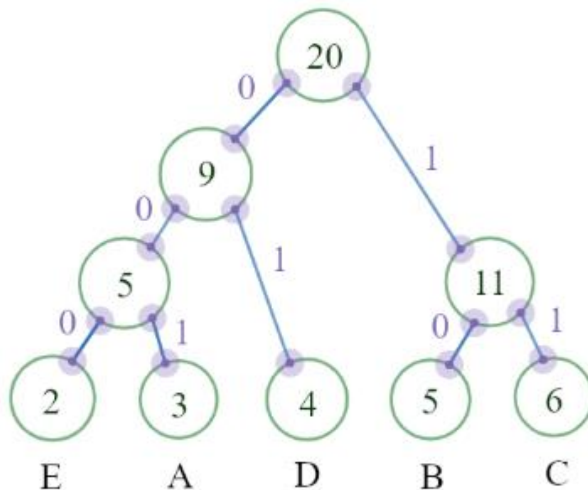
A	B	C	D	E
65	66	67	68	69
01000001	01000010	01000011	01000100	01000101
3	5	6	4	2
3/20	5/20	6/20	4/20	2/20

У четвртом реду табеле је број појављивања појединог слова у датој поруци, а у последњем петом наведене су фреквенције слова, њихове статистичке вероватноће.

Када компјутер преноси ову поруку, аски кодирану, свако слово троши 8 бита (знакова 0 или 1), тако да укупна њена дужина износи $8 \times 20 = 160$ бита. При томе у меморији је негде и списак кодова, што за ових пет актуелних слова износи додатних $8 \times 5 = 40$, односно укупно $160 + 40 = 200$ бита. Већи број бита значи веће трошење меморије носиоца и дуже трајање преноса података.

Међутим, није потребно заузимати по осам позиција појединог слова тако скромне поруке. Самих пет тих слова енглеског алфабета једнако тачно кодирају и три бинарне позиције, рецимо редом: 000, 001, 010, 011, 100, 101. Тада трошимо $3 \times 20 = 60$ бита за текст и још $3 \times 5 = 15$ бита за табелу декодирања, што је додатних $60 + 15 = 75$ бита, евентуално плус $8 \times 5 = 40$ за додефинисање аски ознака, што је укупно $75 + 40 = 115$ бита. Даље згушњавање омогућава променљива дужина кода.

Редамо дата слова по растућој фреквенцији у датој поруци и формирамо граф. Слова су дно графа.



Заокружене изнад слова су фреквенције. Црте изнад кругова воде до збирних фреквенција у новим круговима, све до врха пирамиде. Тако се 5 пута јавља неко слово из скупа {E, A}, или 11 пута из скупа {B, C}, а 9 пута из ширег скупа {E, A, D}, или 20 пута из још ширег скупа свих слова поруке {E, A, D, B, C}.

Једноставно даље, путеви од поруке до слова казују кôд слова. Лево скретање увек има кôд 0, десно 1. Тако се од 20 до E стиже путем 000, тј. слово E прве азбуке кодира 000 друге. Тако се слово A кодира са 001 друге азбуке, слов D се кодира са 01, B са 10, а C са 11.

Цела порука је кодирана са: 101111001101001010010001111101000100001011111. Декодирање је такође једноставно. Идемо од 20 графом путем бројева кодиране поруке. На почетку је „10...“, што значи десно-лево, а то је слово B. Затим опет крећемо од 20 путем „...11...“, десно-десно, што нас доводи до слова C. Након та четири бита имамо опет C и тако даље низ поруку.

[24] 3.4 Huffman Coding - Greedy Method, https://youtu.be/co4_ahEDCho

87. Декод канала

Рекли смо [83.] да је кодирање намерна трансформација извора у облик из којег је могуће опет добити тачан полазни податак. При томе разликујемо заштиту података од неовлаштених особа, податке у рачунару, телеграфију. У наставку узимамо у обзир да је кодирана порука подвргнута сметњама при проласку кроз канал које поруку изобличавају и отежавају декодирање.

Пуна носивост канала је његов тотални шум. То је максимална информација датог канала у некој јединици времена, односно јединичне дужине текста. Оно што се каналом преноси као „користан терет“ је информација мања од те, а теорија предвиђа да су то смисленије поруке што су мање информативне. Такође, то су ефикаснији (брже преносиви) кодови што су мање информативни, као и што су мање информативне оне заједнице датих јединки које би биле боље организоване (ефикасније).

Како то, недавно ме је неко упитао, да виталност живих бића долази од веће информације коју она имају у односу на неживу твар, а онда разумевањем смањује се информација?

Звучи чудно, али не више од апсурдног подизања терета да би се у спуштању искористила његова потенцијална енергија и добио користан рад. Биће без вишка „виталности“ не може „спуштањем“ доћи до корисног „разумевања“, јер за мртву твар важи принцип најмањег дејства, тј. мртва је већ на „дну дна“, са информацијом је у минимуму.

Аналогно енергији, за информацију важи закон одржања. Заправо, из закона одржања укупне информације може се извести закон одржања укупне енергије датог система. Рецимо, за макро системе којима време једнако тече, из константности информације и еквивалентности информације и физичког дејства

$$[\text{информација}] = [\text{енергија}] \times [\text{време}]$$

мерене у константним временским интервалима следи константност енергије.

Наравно, ово је део моје теорије информације, не званичне, а „информација“ у формули је само од оних информација уопште које се (у пакету) крећу физичким простором. Са друге стране, скоро потпуни шум канала могуће је добити и кодирањем, маскирањем, али уз већу цену и мање ефикасним (споријим) кодом.

У овој теорији, брзина протицања времена може се дефинисати и „количином исхода“ тако да где време протиче спорије (у релативном кретању, у јачем гравитационом пољу) постоји и релативни дефицит информације, успоравање реализација. То је једна од старих основних теза (моје) теорије информације. Али њој доследан и још занимљивији је феномен простора веће брзине протицања времена са становишта мање. Тамо већу неизвесност и стога мање вероватноће појединих исхода, убрзане видимо као догађаје чије су шансе повећане. Природа као да равна те две крајности, или еквивалентно као да онима повећава вероватноће и смањује информацију. Начелно, биће да се и на тај начин остварује „спонтана тежња ка мањој емисији информације“.

Помињем те ситуације да лакше разумемо смањивање информације убрзавањем преноса кодних порука, односно да приметимо доследност у том на први поглед „чудном“ резултату. Убрзавања која следе из ефикасности, дакле смањења сувишних опција, иначе су по овој теорији испраћена мањом средњом (Шеноновом) информацијом.

87.1. Брзина

Изоставимо ли захтев за јединственошћу декода, можгли бисмо рећи да је пренос података композиција правог кода и кода канала. Извор података $x_0, x_1, x_2, \dots \in \mathcal{A} = \{a_1, a_2, \dots, a_l\}$ преводимо у извор $y_0, y_1, y_2, \dots \in \mathcal{B} = \{b_1, b_2, \dots, b_r\}$ инјективном функцијом $f: [\![x_i]\!] \to [\![y_j]\!]$ коју зовемо (прави) кôд, ако на основу низа Y-а можемо добити тачан полазни низ X-ова:

$$\text{извор } X \underset{j}{\longrightarrow} Y \longrightarrow \begin{matrix}\text{канал}\\ \text{сметње}\end{matrix} \longrightarrow Z \underset{g}{\longrightarrow} Y' \underset{f^{-1}}{\longrightarrow} X'$$

Кодирани низ Y-а даље пролази кроз канал трпећи промене због сметњи $\widehat{K}: \vec{y} \to \vec{z}$, чиме чиме се део података неповратно губи (постаје непрепознатљив), али је линеарна трансформација $\widehat{K}$ такође нека инјекција.

Најбољом инверзијом $g: z' \to y'$ добија се само приближан улазни низ Y-а, тако да је тачан декод излазне поруке Z немогућ без обзира што функција кода (f) има инверзну функцију декода (f^{-1}). У општем случају се ради о трансформацијама низова са више од једног члана, па се конкретно f и $\widehat{K}$ називају „блок-кодови“.

За разлику од тачног декодирања (f^{-1}), које смо могли постићи у претходном поглављу, сада тражимо најприближније (g) декодирање податка који је прошао кроз канал, које се зове „поступак препознавања“, или „шема одлучивања“. Слично претходним разматрањима и овде нам је важна брзина преноса. Разликујемо пре свега брзину преноса кодних слова и брзину преноса информације.

Када је m-ти члан низа x-ова кодиран у низ y-а укупни број кодираних порука је M. Ако су ове поруке једнако вероватне, њихова укупна информација је $\log M$, па низ y-а по кодном слову садржи информацију

$$R = \frac{\log M}{n}.$$

То је максимална количина информације једног слова n-чланог низа $\vec{y}_j = (y_{j_1}, \dots, y_{j_n})$, са y-има из друге, кодне азбуке ($\mathcal{B}$). Број R се обично зове коефицијент преноса, или брзина блок-кода.

Пример 1. Када би свака комбинација од r кодних слова (азбуке $\mathcal{B}$) давала по једну смислену поруку, тада би било $M = r^n$, тј. $R = \log r$. Ово је максимална информација скупа од r разних елемената, али и максимална брзина датог блок-кода. У општем случају је наравно $R \le \log r$. ❑

Други пример је ергодички извор података (84.2) са фиксном дужином кодних замена, где важи тамо наведена (прва од) неједнакост, тј. $M \le r^n$, или $\log M \le n \log r$, па је $R \le \log r$. Ово R интерпретирамо исто као горе, тј. M је број порука извора које су кодиране у n-члани низ. То је довољан и потребан услов за максималну поузданост кодне азбуке.

Прецизније, у наслову „Потпуност“ [84.] је било

$$2^{m(S-\varepsilon)} \le M \le 2^{m(S+\varepsilon)},$$

где $\varepsilon \to 0$ кад $m \to \infty$, тј. $m(S-\varepsilon) \le \log M \le m(S+\varepsilon)$, $\varepsilon > 0$, или

$$\frac{m}{n} \cdot (S - \varepsilon) \leq R \leq \frac{m}{n} \cdot (S + \varepsilon),$$

где је $m/n = v$ брзина преноса кода по слову извора. За $m \to \infty$ биће $R = vS$.

Са друге стране, ако извор кодирамо променљивом дужином кодних знакова, тада је очекивана вредност дужине елементарне кодне речи $\bar{n}$, тј. средњи број потрошених симбола за једно слово кодиране азбуке. Дефинишемо $V = 1/\bar{n}$, па претходна једнакост даје $1/V = \bar{n} = S/R$. Упоређујући исти резултат са Шенон-Фановом теоремом [85. Теорема 2] добијамо да је $R = \log r$ за оптималан код и даље $v \leq V$. Интуитивно је сасвим јасно да оптималан код има минималну средњу дужину кодне речи, тј. да тај има максималну брзину преноса. Последње једнакости само потврђују да је максимална (интуитивна) брзина достигнута у скупу тренутних кодова.

Интуитивно можемо разумети и релацију $R \leq C$, где је C капацитет канала. Највећа информација извора (C) која се може препознати на излазу датог канала није мања од (R) највеће информације слова y_j. Извор кроз канал иде слово по слово као случајна променљива дата улазном, односно излазном расподелом. Према томе, услов $R > C$ значи да не можемо конструисати кодер и декодер датог канала, без обзира каквом техничком опремом располагали.

87.2. Правило препознавања

Пролазећи кроз канал податку се дешавају промене због којих излазни низ порука неће бити препознат као улазни. То је основни разлог зашто линеарна трансформација $\hat{K}$, коју зовемо канал, није прави кôд. Наиме, ако постоји јединствен начин да се од излаза добије улаз онда је порука прошла непромењена, што у општем случају није тачно:

$$\text{извор } \vec{X} \underset{\text{сметње}}{\overset{\hat{K}}{\dashrightarrow}} \text{ пријем } \vec{Y} \underset{g}{\dashrightarrow} \vec{X}'$$

Функција g није прави декод јер није јединствена. Зато јој додељујемо посебан назив.

Дефиниција 1. Правило препознавања (или декод канала) је свака функција $g: \vec{y} \to \vec{x}'$, где су $\vec{y}$ и $\vec{x}'$ низови слова. ▨

Међу свим „правилима препознавања“ (g) нама су најинтересантнија она која дају $\vec{x}'$ најприближније вредностима $\vec{x}$.

Дефиниција 2. (Правило препознавања најмањом вероватноћом грешке) Функција $\vec{x}' = g(\vec{y})$ је дефинисана једнакошћу $\Pr(\vec{x}'|\vec{y}) = \max_{\vec{x}} \Pr(\vec{x}|\vec{y})$. ▨

Полазећи од дефиниције (2), због $\Pr(\vec{x}|\vec{y}) = \Pr(\vec{x},\vec{y})/\Pr(\vec{y})$, добијамо еквивалентну једнакост $\Pr(\vec{x}'|\vec{y}) = \max_{\vec{x}} \Pr(\vec{x},\vec{y})$. Познато је да у општем случају важи:

$$\sum_{\vec{x}}\sum_{\vec{y}} \Pr(\vec{x},\vec{y}) = 1, \quad \sum_{\vec{x}} \Pr(\vec{x}) = 1, \quad \sum_{\vec{y}} \Pr(\vec{y}) = 1,$$

$$\Pr(\vec{y}) = \sum_{\vec{x}} \Pr(\vec{x},\vec{y}), \quad \Pr(\vec{x},\vec{y}) = \Pr(\vec{x})\Pr(\vec{y}|\vec{x}),$$

где је $\vec{y}$ реч од $n = 1, 2, 3, \ldots$ слова азбуке $\mathcal{B} = \{b_1, \ldots, b_r\}$, док на улазу имамо реч $\vec{x}$ од $m = 1, 2, 3, \ldots$ слова азбуке $\mathcal{A} = \{a_1, \ldots, a_l\}$. У сваком случају овде имамо збирове са коначно много сабирака.

Када за свако дато $\vec{y}$ нађемо одговарајуће $\vec{x}'$ такво да је $\Pr(\vec{x}', \vec{y}) \geq \Pr(\vec{x}, \vec{y})$ за сваку улазну поруку $\vec{x}$ онда имамо дефинисано правило препознавања (g) које даје најмању вероватноћу грешке. Наиме, тада је $1 - \Pr(\vec{x}, \vec{y}) \geq 1 - \Pr(\vec{x}'|\vec{y})$ за свако $\vec{x}$, па је тотална вероватноћа грешке (Error, за правило препознавања дефиниције 2):

$$P_2(E) = \sum_{\vec{y}} \Pr(\vec{y})[1 - \Pr(\vec{x}'|\vec{y})]$$

минимална.

Због последње једнакости задато правило препознавања је најинтересантније. Највећи недостатак овог правила је потреба познавања свих вероватноћа $\Pr(\vec{x})$ што је често практично немогуће. Са друге стране, познавање свих вероватноћа $\Pr(\vec{x})$ чини функцију g непотребном.

Дефиниција 3. (Правило идеалног посматрача) Функција $\vec{x}' = g(\vec{y})$ је задата једнакошћу

$$\Pr(\vec{y}|\vec{x}') = \max_{\vec{x}} \Pr(\vec{y}|\vec{x})$$

. ▨

Практично, излазу $\vec{y}$ додељујемо онај улаз $\vec{x}'$ за који је тај излаз највероватнији. Ако се ради о једночланим низовима, улаз и излаз, онда имамо матрицу канала $\hat{K}$ чији је члан $k_{ij} = \Pr(y|x)$, где x је i-то слово азбуке $\mathcal{A} = \{a_1, \dots, a_l\}$ а y је j-то слово азбуке $\mathcal{B} = \{b_1, \dots, b_r\}$. Матрица је реда $l \times r$.

Пример 2. Канал је дат матрицом

$$\hat{K} = \begin{pmatrix} 0{,}6 & 0{,}1 & 0{,}3 \\ 0{,}5 & 0{,}3 & 0{,}2 \\ 0{,}4 & 0{,}2 & 0{,}4 \end{pmatrix}$$

одакле видимо да је улазна азбука $\mathcal{A} = \{a_1, \dots, a_3\}$, а излазна $\mathcal{B} = \{b_1, \dots, b_3\}$, тј. $l = r = 3$. Правило препознавања (деф. 3) у свакој колони $j = 1, 2, 3$ издваја највећи члан k_{ij} па добијамо:

$$a_1 = g(b_1), \quad a_2 = g(b_2), \quad a_3 = g(b_3).$$

Према томе, излазни низ $y_1, y_2, y_3, \dots, y_j \in \mathcal{B}$ препознаје се тако да се свако b_j протумачи као a_j. ❑

У општем случају имамо n-члани низ $\vec{y} = (y_1, \dots, y_n)$ и неки скуп Y таквих n-чланих низова, па је условна вероватноћа

$$\Pr(Y|\vec{x}) = \sum_{\vec{y} \in Y} \Pr(\vec{y}|\vec{x})$$

да излаз $\vec{y}$ припада скупу Y који је опет, подскуп свих речи дужине n над азбуком $\mathcal{B}$. Ако $\Pr(Y|\vec{x})$ је вероватноћа тачног преноса података $\vec{x}$, тада је $1 - \Pr(Y|\vec{x})$ вероватноћа грешке, па је

$$P_3(E) = \sum_{\vec{x}} \Pr(\vec{x}) \cdot [1 - \Pr(Y|\vec{x})]$$

тотална вероватноћа грешке. Доказаћемо да су тоталне вероватноће грешака за два правила препознавања једнаке.

Ако уређен пар $(\vec{x}, \vec{y})$, тј. улаз-излаз канала, припада неком од скупова

$$T_x = \{(\vec{x}, \vec{y}) : \vec{x} = g(\vec{y})\}$$

онда је пренос кроз канал без грешке, док у супротном случају имамо грешку при препознавању правилом g. Зато је тотална вероватноћа грешке према 3. дефиницији:

$$P(E) = \sum_{\vec{x}} \sum_{\vec{y} \notin Y_x} \Pr(\vec{x}, \vec{y}) = \sum_{\vec{x}} \Pr(\vec{x}) \sum_{\vec{y} \notin Y_x} \Pr(\vec{y}|\vec{x}) =$$

$$= \sum_{\vec{x}} \Pr(\vec{x}) \cdot [1 - \Pr(Y_x|\vec{x})] = P_3(E),$$

где је кориштена ознака $Y_x = \{\vec{y} : (\vec{x}, \vec{y}) \in T_x\}$.

Према 2. дефиницији имамо:

$$P(E) = \sum_{\vec{y}} \sum_{\vec{x} \neq g(\vec{y})} \Pr(\vec{x}, \vec{y}) = \sum_{\vec{y}} \Pr(\vec{y}) \sum_{\vec{x} \neq g(\vec{y})} \Pr(\vec{x}|\vec{y}) =$$

$$= \sum_{\vec{y}} \Pr(\vec{y}) \cdot [1 - \Pr(\vec{x}|\vec{y})] = P_2(E).$$

Дакле, доказали смо да је $P_2(E) = P_3(E)$.

Отуда су поменуте предности правила препознавања (2) уједно и предности правила (3), с тим да правило (3) можемо заиста употребити ако познајемо матрицу канала. Недостатак ових правила је у чињеници да тотална вероватноћа грешке може бити најмања и онда када се поједини сигнал сасвим погрешно дешифрује.

Пример 3. Канал је дат матрицом

$$\hat{K} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0,5 & 0,5 & 0 \end{pmatrix}$$

Видимо да су највећи чланови прве и друге колоне у доњој врсти, док је у трећој колони свеједно узимамо ли горњи или средњи члан. Правило препознавања (3) је одређено са

$$g(b_1) = a_3, \quad g(b_2) = a_3, \quad g(b_3) = a_1.$$

Не сумњамо да је тотална вероватноћа грешке овако најмања, али видимо да ће нпр. сигнал a_2 бити увек погрешно препознат, јер је вероватноћа грешке при преносу сигнала a_2 једнака јединици. ❑

У следећем наслову размотрићемо неке компромисе између тоталне и појединачне вероватноће грешке, у том смислу да правило препознавања не направи велику грешку при препознавању појединих сигнала, али да уз такво ограничење добијемо што је могуће мање $P(E)$.

87.3. Блок-кôд

Поред циља постављеног на крају претходног поднаслова, уводимо релације међу основне појмове као што су: брзина преноса (R), капацитет канала (C) и правило препознавања (g) за речи дужине n слова датог извора, које представљају фундаменталне резултате теорије декодирања канала. Овде се пролазак података кроз канал, односно линеарни оператор $\hat{K}$ третира као кôд чији је декод правило декодирања.

Проласком слова $x_i \in \mathcal{A} = \{a_1, \dots, a_l\}$ кроз канал добијамо слово $y_j \in \mathcal{B} = \{b_1, \dots, b_r\}$, а проласком једне од M речи $\vec{x} = (x_1, \dots, x_n)$ добијамо реч $\vec{y} = (y_1, \dots, y_n)$. Свака од M речи улаза трансформише се у једну од речи излаза формирајући највише N парова $(\vec{x}, \vec{y})$, који се често називају елементима блок-кода.

Ако је $\vec{x}$ улазни, а $\vec{y}$ излазни низ слова датог канала $\widehat{K}$, тада је

$$I(\vec{x}, \vec{y}) = \log \frac{\Pr(\vec{x}, \vec{y})}{\Pr(\vec{x}) \cdot \Pr(\vec{y})}$$

узајамна информација елемената блок-кода $(\vec{x}, \vec{y})$, где су $\Pr(\vec{x})$ и $\Pr(\vec{y})$ вероватноће појаве речи $\vec{x}$ и $\vec{y}$ редом на улазу и излазу из канала. Специјално, ако извор нема меморију, тј. ако је:

$$\Pr(\vec{x}) = \Pr(x_1, \dots, x_n) = \Pr(x_1) \dots \Pr(x_n)$$

биће:

$$I(\vec{x}, \vec{y}) = \log \frac{\Pr(\vec{y} | \vec{x})}{\Pr(\vec{y})} = \sum_{i=1}^{n} \log \frac{\Pr(y_i | x_i)}{\Pr(y_i)} = \sum_{i=1}^{n} I(\vec{x}, \vec{y}).$$

Према томе, узајамна информација пара n-чланих речи једнака је збиру узајамних информација одговарајућих појединих парова слова, ако су слова у речи међусобно независна.

Када извор информације има меморију, а чланови матрице канала су константе, можемо посматрати $I_i = I(x_i, y_i)$, $i = 1, \dots, n$, као низ независних променљивих, јер канал са константним коефицијентима нема меморију. Математичко очекивање таквог низа је:

$$EI_i = \sum_{i} \Pr(x_i, y_i) \cdot I_i = I(\vec{p}, \vec{q}).$$

Дакле, једнако је средњој узајамној информацији датог канала, где је $a_i \in \mathcal{A}$, $p_i = \Pr(a_i)$, $b_j \in \mathcal{B}$, $q_j = \Pr(b_j)$, док је $\vec{p} = (p_1, \dots, p_l)$, $\vec{q} = (q_1, \dots, q_r)$. Ако је $\vec{p}$ она улазна расподела која даје максимално $I(\vec{p}, \vec{q})$ онда је $EI_i = C$, тј. очекивање је једнако капацитету канала.

У општем случају, ако канал преноси информацију слово по слово, такође има смисла дефинисати збир информација појединих слова

$$I = \sum_{i=1}^{n} I_i.$$

Због претходне једнакости, тада је у крајњем случају $EI = nC$. Али за разлику од ње, која даје средњу информацију по слову, ова је збир информација од n слова.

Случајни низ независних бројева I_i има информацију S која је гранична вредност од

$$S_n = -\frac{1}{n} \sum_{i=1}^{n} \Pr(I_i) \log_b \Pr(I_i),$$

за $n \to \infty$, те је $N \sim b^{nS}$. Због $N \geq M$ даље добијамо $S \geq R$. Дакле, неодређеност појаве пара улаз-излаз унапред задате узајамне информације није мања од информације по једном слову (R). Слично добијамо $N \geq b^{nR}$, или $C \geq R$, где важи једнакост ако и само ако $N = M$, тј. ако свака комбинација слова, како улазне тако и излазне азбуке, чини реч и када је број речи улазне азбуке једнак броју речи излазне.

Овде су нам мање интересантни парови $(\vec{x}, \vec{y})$ са мало узајамне информације и, са друге стране, парови који се реализују са вероватноћом нула. Ради тога делимо све улазно-излазне речи у два скупа:

$$A' = \{(\vec{x}, \vec{y}) : I(\vec{x}, \vec{y}) > a\}, \quad A" = \{(\vec{x}, \vec{y}) : I(\vec{x}, \vec{y}) \leq a\},$$

где је a произвољан али фиксан реалан број ($a \in \mathbb{R}$). Како су A' и $A"$ дисјунктни скупови са N' и $N"$ елемената респективно, то је $N = N' + N"$. Поредамо ли све улазне речи у низ $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_M$ па узмемо произвољан елеменат $\vec{x}$ тога низа, можемо дефинисати скуп излазних речи

$$A'_x = \{\vec{y} : (\vec{x}, \vec{y}) \in A'\}.$$

Да бисмо елиминисали пар $(\vec{x}, \vec{y})$ који се реализује са вероватноћом нула узмимо неки мали број $\varepsilon > 0$, па из скупова $A'_{x_1}, \dots, A'_{x_M}$ избацимо елементе са излазном вероватноћом мањом од ε, затим избацимо елементе A'_{x_i}-тог скупа који већ постоје у претходном $i = 1, 2, \dots, M$. Добијамо низ дисјунктних скупова ($i = 1, 2, \dots, M$):

$$S'_1 = \{\vec{y} : (\vec{x}_1, \vec{y}) \in A'\}, \quad S'_i = \{\vec{y} : (\vec{x}_i, \vec{y}) \in A'\} \setminus \bigcup_{j=1}^{i-1} S'_j, \quad \Pr(S'_i | \vec{x}_i) \geq \varepsilon.$$

Лако је видети да је $S'_i \cap S'_j = \emptyset$ кад $i \neq j$, док скуп $S' = \bigcup_{k=1}^{M} S'_k$ садржи све парове скупа A'.

Означимо са $S"$ скуп комплементаран скупу S', дакле скуп маловероватних парова, па приметимо да је

$$\sum_{Y \in S" \cap A'_x} \Pr\,(\vec{y}|\vec{x}) < \varepsilon,$$

за свако $\vec{x}$. Са друге стране је:

$$\sum_{y \in S' \cap A'_x} \Pr\,(\vec{y}|\vec{x}) \leq \sum_{y \in S'} \Pr(\vec{y}|\vec{x}) = \Pr(S'|\vec{x}).$$

Уведимо ознаку $\Pr(A') = \sum_{(\vec{x},\vec{y}) \in A'} \Pr\,(\vec{x}, \vec{y})$. По дефиницији условне вероватноће је даље:

$$\Pr(A') = \sum_{\vec{x}} \Pr\,(\vec{x}) \sum_{Y \in A'_x} \Pr\,(\vec{y}|\vec{x}) = \sum_{\vec{x}} \Pr\,(\vec{x}) \sum_{\vec{y} \in S' \cap A'_x} \Pr\,(\vec{y}|\vec{x}) + \sum_{\vec{x}} \Pr\,(\vec{x}) \sum_{\vec{y} \in S" \cap A'_x} \Pr\,(\vec{y}|\vec{x}),$$

где се сабира по свим улазним речима $\vec{x}$. Последњи сабирак, због $\sum_{\vec{x}} \Pr(\vec{x}) = 1$ и због горње строге неједнакости, није већи од ε. Претходни сабирак, због претходне неједнакости, постаје:

$$\sum_{\vec{x}} \Pr(\vec{x}) \Pr\,(S'|\vec{x}) \leq \Pr(S') = \sum_i \Pr(S'_i) \leq \sum_i \Pr\big(A'_{x_i}\big),$$

због $S'_i \subseteq A_{x_i}$. По дефиницији (2), за скуп A' добијамо $\Pr(\vec{y}|\vec{x}) > b^a \Pr(\vec{y})$, па је:

$$\Pr(A'_{x_i}) = \sum_{\vec{y}\in A_{x_i}} \Pr(\vec{y})\, b^{-a} \cdot \sum_{\vec{y}\in A_{x_i}} \Pr(\vec{y}|\vec{x}) = b^{-a} \Pr(A'_{x_i}|\vec{x}) \leq b^{-a},$$

где је $b > 1$ база логаритма информације. Према томе, претходни сабирак није већи од Mb^{-a}, а оба заједно дају:

$$\Pr(A') \leq Mb^{-a} + \varepsilon, \quad \text{тј.} \quad M \geq b^a[\Pr(A') - \varepsilon].$$

Последњи резултат можемо интерпретирати као доњу границу броја (M) улазних речи у дати канал за које се може конструисати правило препознавања $\vec{x}_i = g(\vec{y})$, $\vec{y} \in S'_i$. За разлику од претходних (87.2) ово правило има појединачну вероватноћу грешке већу од $1 - \varepsilon$. Пратећи доказ, видимо да последње неједнакости важе и за оне канале који имају меморију.

Међутим, за веће ε могу бити сви скупови S'_i $(i = 1, 2, \dots)$ празни, па од правила препознавања (g) нема ништа. Исто тако, независно од ε, али са великим бројем a можемо добити премален па чак и празан скуп A', губећи речи изворног језика. У оба случаја доња граница броја M у последњој релацији је тако малена да број (M) дешифрованих речи постаје недовољан.

Како изворних речи дужине n слова има $M = b^{nR}$, где је R коефицијент брзине, то је основно питање декодирања канала: може ли десна страна те неједнакости бити једнака броју свих речи $\vec{x} = (x_1, \dots, x_n)$? Одговор на то питањање даје следећа, тзв. фундаментална теорема.

Теорема 1. За дискретни канал без меморије, капацитета $C > 0$ и за реални број R $(0 < R < C)$ постоји блок-код од M речи са по n слова и правило препознавања g тако да је:

$$\varepsilon = \min_{1\leq i\leq M} \Pr(S_i|\vec{x}) \geq 1 - \alpha \cdot e^{-\beta n}, \quad M \geq b^{nR},$$

где параметри $\alpha > 0$ и $\beta > 0$ не зависе од n.

Доказ: Када је z позитивна случајна променљива са позитивном дисперзијом логаритам дефинишемо очекивања:

$$\eta(t) = \ln E(e^{tz}), \quad t \in \mathbb{R},$$

$$\frac{d^2\eta(t)}{dt^2} = \frac{E(z^2 e^{tz}) \cdot E(e^{tz}) - (E(z \cdot e^{tz}))^2}{(E(e^{tz}))^2}.$$

Шварцова неједнакост за $z_1 = e^{tz/2}$ и $z_2 = ze^{tz/2}$ даје:

$$(E(z_1 \cdot z_2))^2 \leq E(z_1^2) \cdot E(z_2^2),$$

$$(E(ze^{tz}))^2 \leq E(e^{tz}) \cdot E(z^2 e^{tz}),$$

$$\eta''(t) > 0.$$

Када би овде важила једнакост биле би z_1, z_2 линеарно зависне случајне променљиве за које је:

$$\alpha_1 \cdot z_1 + \alpha_2 \cdot z_2 = (\alpha_1 + \alpha_2 \cdot z) \cdot e^{\frac{tz}{2}} = 0.$$

Услов $\alpha_2 \neq 0$ значи да је $z = -\alpha_1/\alpha_2$ те да z нема претпостављену позитивну дисперзију. Ставимо ли $\xi(t) = \eta(t) - t \cdot \eta'(t)$ добијамо $\xi'(t) = -t \cdot \eta''(t)$. За $t > 0$ је $\xi'(t) < 0$ па је тада $\xi(t)$ опадајућа функција, а за $t < 0$ је $\xi(t)$ растућа. Зато $\xi(0) = \max_{t<0} \xi(t)$.

Отуда једнакости $\xi(0) = \eta(0) = \ln E(1) = 0$ дају $\eta(t) - t \cdot \eta'(t) < 0$ за $t < 0$. Сменом (Марковљева неједнакост) $z_0 = e^{Cz}$ добијамо:

$$\Pr\,(z_0 \geq e^{\alpha C} \cdot E(z_0),$$

$$\Pr(e^{Cz} \geq e^{\alpha C}) \leq e^{-\alpha C} \cdot E(e^{Cz}),$$

$$\Pr(z \geq \alpha) \leq e^{-\alpha C} \cdot E(e^{Cz}).$$

Сменом $z \to -I$, $\alpha \to -\alpha$ имамо даље:

$$\Pr(-I \geq -\alpha) = \Pr(I \leq \alpha) \leq e^{\alpha C} \cdot E(e^{-CI}), \quad C > 0.$$

Користећи претходну формулу за збирну информацију појединих слова и овај резултат Шварцове неједнакости, налазимо:

$$E(e^{-CI}) = E\left(e^{-C\sum_{i=1}^{n} I_i}\right) = \prod_{i=1}^{n} E(e^{-CI_i}) = \left(e^{\eta(-C)}\right)^n.$$

Ставимо ли $t = -C < 0$ биће $E(e^{tI}) = e^{n \cdot \eta(t)}$, $t < 0$, па је $(z = I_i)$ $\eta'(0) = E(I_i) = C$. Та неједнакост $(\eta''(t) > 0)$, дакле, показује да $\eta(t)$ строго расте на целом скупу реалних бројевав и да постоји број $t_0 < 0$ такав да:

$$\frac{C+R}{2} < \eta'(t_0) < C,$$

$$\Pr\left(\frac{I}{n} \leq \frac{C+R}{2}\right) \leq \Pr\left(\frac{I}{n} \leq \eta'(t_0)\right) \leq e^{n\left(\eta(t_0) - t_0\eta'(t_0)\right)}.$$

Посебно је $\Pr\left(I \leq n \cdot \frac{C+R}{2}\right) < e^{n \cdot \xi(t_0)}$, $\Pr(A'') \leq e^{n\beta}$, где је стављено $-\beta = \eta(t_0) - t_0\eta'(t_0)$, а у познатој релацији $\Pr(A^c) = 1 - \Pr\,(A)$, комплемент $A^c = A''$, $A = A'$. Због $\eta(t) - t \cdot \eta'(t) < 0$ биће $\beta > 0$, што значи да $\Pr\,(A'') \to 0$ када $n \to \infty$. Другим речима, постоји довољно велико n тако да је $\delta = e^{-n \cdot \frac{C-R}{2}} + e^{-n \cdot \beta} < 1$. При томе, сменом $\delta = 1 - \varepsilon$, $b = e$ у $M \geq b^a[\Pr(A') - \varepsilon]$ и $n \cdot \frac{C+R}{2} = \alpha$ добијамо $M \geq e^{-n \cdot \frac{C-R}{2}} \cdot e^{n \cdot \frac{C+R}{2}} = e^{nR}$, те $1 - \varepsilon < 2 \cdot e^{n \cdot \min\left\{\frac{C-R}{2} \cdot \xi_0\right\}}$, а отуда и тврђење теореме. ■

У литератури, ову „фундаменталну теорему“ налазимо и као $\varepsilon = \max_{1 \leq i \leq M} [1 - \Pr\,(S_i|x_i] \leq \alpha \cdot e^{-\beta n}$, у овим ознакама и условима, што је њен еквивалентан облик.

87.4. Немогућ захтев

Видели смо доказ познате „фундаменталне теореме“ о декодирању канала. Постоји и обрат ове теореме (Wolfowitz) такође потежак, коју ћу приказати у мало скраћеној форми. Основна теорема потврђује егзистенцију кодера и декодера дискретног канала без меморије, брзине R не веће од капацитета C канала и произвољно великом потпуношћу, тј. произвољно малом вероватноћом ε

максималне грешке преноса. Обратна теорема тврди да је то немогуће постићи са брзином преноса већом од капацитета канала.

Знамо да више неизвесности имамо погађајући једно слово изван текста, него слова у тексту. То је доказивано овде [80.1. Теорема 1] релацијама $s_{n+1} \le s_n$, $S_{n+1} \le S_n$, $n = 1, 2, 3, \dots$, где је s_n информација коју има n-то слово у низу када је познато претходних $n-1$, а S_n је средња својствена информација једног слова у низу од n слова. Поменуте релације постају једнакости ако и само ако су слова у тексту међусобно независна.

За разлику од ових једнакости, оне из претходног поднаслова [87.1. Брзина] односе се на узајамне информације парова слова блок-кода. Међутим, већ смо доказали да је трећа једнака првој, ако су слова у речи извора међусобно независна.

Нека је $\vec{x} = (x_1, \dots, x_n)$ одређена улазна реч у дати канал, а $\vec{y} = (y_1, \dots, y_n)$ произвољна могућа одговарајућа излазна реч, где $x_i \in \mathcal{A}$, $y_j \in \mathcal{B}$. Тада је условна информација излаза уз познати улаз

$$S(\vec{y}|\vec{x}) = -\sum_{\vec{y}} \Pr(\vec{y}|\vec{x}) \log \Pr(\vec{y}|\vec{x}),$$

где се сабира по свим излазним речима $(\vec{y})$ дужине n слова. Без обзира на меморију извора, ако канал нема меморију биће:

$$\Pr(\vec{y}|\vec{x}) = \Pr(y_1|x_1) \cdots \Pr(y_n|x_n),$$

$$S(\vec{y}|\vec{x}) = -\sum_{\vec{y}} \Pr(y_1|x_1) \cdots \Pr(y_n|x_n) \left[\log \Pr(y_1|x_1) + \cdots + \log \Pr(y_n|x_n)\right],$$

$$S(\vec{y}|\vec{x}) = \sum_{i-1}^{n} S(y_i|x_i).$$

Условна информација излазне речи једнака је збиру условних информација појединих слова, за дискретни канал без меморије. Посматрамо ли $\vec{y}$ као произвољни извор, на основу поменутих једнакости, добијамо

$$S(\vec{y}) \le \sum_{i=1}^{n} S(y_i),$$

где вреди једнакости ако су $y_1, \dots, y_n$ независни.

Отуда, због $I(\vec{x}, \vec{y}) = S(\vec{y}) - S(\vec{y}|\vec{x})$ имамо:

$$I(\vec{x}, \vec{y}) \le \sum_{i=1}^{n} [S(y_i) - S(y_i|x_i)] = \sum_{i=1}^{n} I(x_i, y_i).$$

Бирамо ли произвољно $\vec{x}$, ова неједнакост говори о излазној n-чланој речи дискретног канала без меморије. Она постаје једнакост ако и само ако $\Pr(\vec{y}) = \Pr(y_1) \dots \Pr\,(y_n)$ за сваки излаз $\vec{y}$.

Према томе, целокупна информација пренесена каналом није већа од збира појединих узајамних информација. Придружимо ли једну узајамну информацију пара слова (x_i, y_i) блок-кода слову z_i нове азбуке (новог извора) моћи ћемо применити формалне резултате претходних поглавља, где је ово само један од њих. Уместо таквог правца, у наставку завршићемо са разматрањима о декодирању канала.

Означимо са $\vec{y} = (y_1, \dots, y_n)$ одређену излазну реч датог канала, а са $\vec{x} = (x_1, \dots, x_n)$ произвољну улазну, где су $x_i \in \mathcal{A}$ и $y_i \in \mathcal{B}$. Тада је условна информација улаза уз познати излаз

$$S(\vec{y}|\vec{x}) = -\sum_{\vec{x}} \Pr(\vec{x}|\vec{y}) \log \Pr(\vec{x}|\vec{y}),$$

где се сабира по свих M различитих улазних речи $\vec{x}$.

Узмимо било које правило препознавања (g) па одредимо једно $\vec{x}' = g(\vec{y})$ из скупа свих улазних речи. Тада је вероватноћа грешке појединачног декодирања:

$$q(\vec{y}) = 1 - \Pr(\vec{x}'|\vec{y}) = \sum_{\vec{x} \neq \vec{x}'} \Pr(\vec{x}|\vec{y}).$$

Због основних особина информације је

$$S(\vec{x}|\vec{y}) = S(q, 1-q) + qS(1) + (1-q)S(q_1, \dots, q_{M-1}),$$

где стоје скраћенице

$$q_i = q_i(\vec{y}) = \frac{\Pr\,(\vec{x}_i|\vec{y})}{1 - q(\vec{y})}, \quad i = 1, \dots, M-1,$$

а $\vec{x}_1, \dots, \vec{x}_{M-1}$ су све n-члане речи извора осим $\vec{x}'$. Максимална информација последњег сабирка је $\log(M-1)$, која одговара скупу једнако вероватних речи извора, док је $S(1) = 0$. Према томе је

$$S(\vec{x}|\vec{y}) \leq S(q, 1-q) + (1-q)\log(M-1).$$

Лева страна достиже максимум када је за примљено $\vec{y}$ једнако вероватна свака од M послатих речи $\vec{x}$. Тада је $S(\vec{x}|\vec{y}) = \log M$, а десна страна ову постиже у оба случаја $q = 1/M$ или $q = 1 - 1/M$. У свим осталим случајевима ова неједнакост је једнакост.

Дефинишемо средњу условну информацију улаза са познатим излазом уобичајено, па је

$$S(\vec{x}|\vec{y}) \leq \sum_{\vec{y}} \Pr(\vec{y}) \left[S(q, 1-q) + \log(M-1) \sum_{\vec{y}} \Pr(\vec{y})\,(1-q) \right].$$

Због особине конвексности информације добијамо:

$$\sum_{\vec{y}} \Pr(\vec{y})\, S(q, 1-q) \leq S\left[\sum_{\vec{y}} \Pr(\vec{y})\, q\,, \sum_{\vec{y}} \Pr\,(\vec{y})(1-q) \right] = S[\Pr(E)\,, 1 - \Pr\,(E)],$$

где је $\Pr(E) = \sum_{\vec{y}} \Pr\,(\vec{y})[1 - q(\vec{y})] = Q_n$ вероватноћа тоталне грешке, па уместо претходног имамо

$$S(\vec{x}|\vec{y}) \leq S(Q_n, 1 - Q_n) + Q_n \log(M - 1).$$

Тиме смо доказали Фанову лему која важи за сваки блок-код (n, M) и правило (g).

Теорема 2. (Обрнута фундаменталној) Ако је за дискретни канал без меморије $R > C$, тада ε не тежи јединици када $n \to \infty$ нити за један блок-код.

Доказ: Претпоставимо да је $\Pr(\vec{x}_i) = 1/M$. Даље је:

$$S(\vec{x}) = -\sum_{i=1}^{M} \Pr(\vec{x}_i) \log \Pr(\vec{x}_i) = \log M,$$

$$I(\vec{x}, \vec{y}) = S(\vec{x}) - S(\vec{x}|\vec{y}) = \log M - S(\vec{x}, \vec{y}),$$

$$\log M - S(Q_n, 1 - Q_n) - Q_n \log(M - 1) \leq I(\vec{x}, \vec{y}),$$

$$[I(\vec{x}, \vec{y}) \leq n - C, \quad S(Q_n, 1 - Q_n) \leq \log 2]$$

$$Q_n > 1 - \frac{\log 2 + nC}{\log M}.$$

Према дефиницији тоталне вероватноће грешке (Q_n) и претпоставке, имамо

$$Q_n = \frac{1}{M} \sum_{i=1}^{M} [1 - \Pr\,(S_i|\vec{x}_i)].$$

По претпоставци теореме, можемо ставити $R = C + \delta$, $\delta > 0$, тј. $M = e^{M/(C+\delta)}$, па је

$$Q_n > 1 - \frac{C + \frac{1}{n}\log 2}{C + \delta} \to 1 - \frac{C}{C + \delta},$$

за $n \to \infty$, што значи да вероватноћа кода (ε) не тежи 1. Према томе, нема потпуног правила препознавања овог канала. Тиме је теорема доказана. ■